

Performance Evaluation of Mobile Network Interconnections in Nigeria

Ohaneme Cletus O., Akpado Kenneth A., Alumona Theophilus L., Emenike Ijeoma S.

Abstract

The adoption of the Interconnect Clearing Houses (ICH) in Nigeria is a long awaited development expected in the telecommunication industry. This paper presents an insight into network performance evaluation and quality of service (QoS) improvement of mobile cellular systems in Nigeria using an ICH as a case study. The components of mechanisms of analyzing and evaluating the various networks are discussed. The paper also identifies the important Key Performance Indicators (KPIs) for QoS evaluation which are used in evaluating the Mobile networks. Two assessment parameters such as Call Completion Ratio (CCR) and Answer Seizure Ratio (ASR) for evaluating the traffic analysis of the networks in Nigeria are deployed. The parameters are applied on four mobile network systems in Nigeria using the general models simulated in Matlab for evaluating network performance. The result of the study shows that the QoS of mobile system in the country is not 100% reliable and still needs to be improved upon.

Keywords: ASR, CCR, Interconnect Clearing House, Mobile network systems, NCC, QoS.

1 Introduction

When Cellular Networks (GSM and CDMA) was introduced in Nigeria in the early 2000's, it was considered an over specified system. But nowadays it is obvious that the whole range of service is widely in use. In the cellular systems, the performance and the quality of service is affected due to the rapid increase in the number of mobile subscribers in recent time.

A cellular network is a radio network distributed over land areas called cells, each served by at least one fixed-location transceiver called cell site or base station. When joined together, these cells provide radio coverage over wide geographic area. This enables a large number of portable transceivers (e.g. mobile phones, pagers, etc) to communicate with each other and with fixed transceivers and telephone anywhere in the network. These large numbers of users can access the network system through the base station transceiver. This is why it is usually referred to as access point. The user migrates across several base

stations thereby allowing the transceivers to transfer communication link to the neighboring base station as it transverses from one area to another.

This paper evaluates the interconnectivity in the telecommunications industry in Nigeria using Interconnect Clearing House Nigeria Limited as a case study. Interconnect clearing house is one out of the six companies licensed in 4th August 2004 by Nigerian Communications Commission (NCC) to provide interconnectivity to all the telecom operators in Nigeria [1]. NCC is the telecommunication service regulatory body in Nigeria. Interconnection refers to the establishment of electronic linkages between service providers so that they can conduct business transactions electronically. Prior to the issuing of these licenses, networks were connected directly to each other. Meaning if there are n networks in a place, there will be $n(n-1)$ links to them as shown in figure 1. With the clearing house, these networks will provide only one link to the clearing

house through which they can reach the other networks as shown in figure 2.

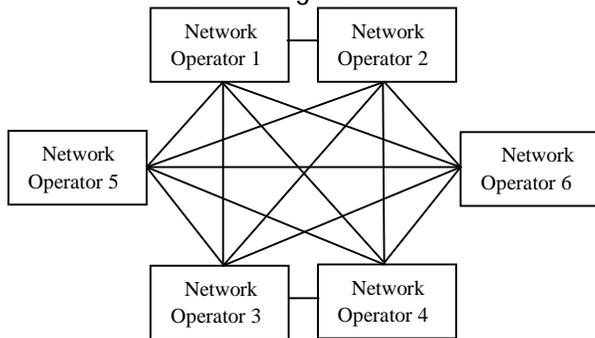


Figure 1: Direct connection of Networks for interconnectivity (i.e., a web)

The clearing house brings in an easy interconnectivity between networks especially the new entrants (new networks). Because to provide telecomm service in Nigeria there must be transmission links to all the respective network operators, which is a whole lot of expenses to start a business. But with clearing house, a link connects the entire networks together in Nigeria.

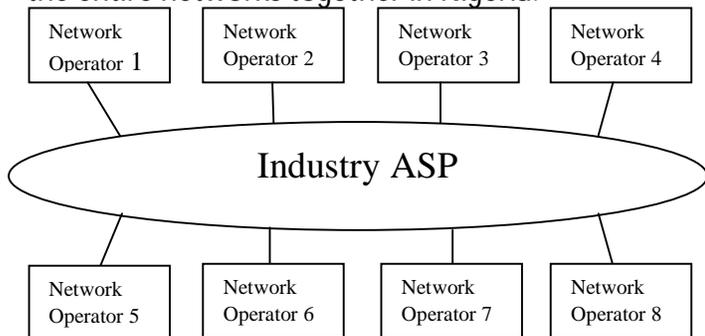


Figure 2: The Clearinghouse solution to interconnectivity

Interconnection is simply an e-commerce or business-to-business trade between and among carriers. Interconnectivity of telecommunication Operators in Nigeria has come a long way with series of challenges from Equipment Inter-Operability of Call Data Record (CDR) for the purpose of settling the Bills of the very traffics terminated. This is to ensure efficient and high quality of services to the end user at a very competitive price since Operators are growing very fast and expanding into all types of telephony – fixed, mobile, Long Distance (National & International) and also into convergent networks[2].

Interconnection is one of the problems that are emerging with the increase in number of operators with open market conditions and Interconnection licensing requirements. This possibly calls for mandatory interconnections between each of Mobile operator, Basic and National/ International Long Distance Operator in any particular licensed service area. Some types of Interconnections in Multi-Operator Multi Service environment as in [3] include: (i) Basic to Basic (Local connection within same community), (ii) Basic to Cellular, (iii) Basic to National/ Long distance, (iv) Basic to International/Long distance (direct or through national long distance operator), (v) Mobile to Mobile, (vi) Mobile to National Long Distance, (vii) Mobile to International Long Distance (direct or through national long distance operator).

2 The general frame work of Interconnection of mobile networks

In the general framework of interconnection of mobile networks, new operators set up electronic connections (called Gateways) to each of the existing operators for exchange of services. In fact, the challenge of interconnection, more than any other factor, limits service providers to doing business on a regional basis with just one or two partners, rather than on a national basis, because of the number of interconnections that may be involved.

To deliver the required nationwide service, a new operator had to establish not only one connection but also multiple connections, (i.e., a separate electronic gateway) to each major mobile operator in Nigeria (See also figure 1). The cost of establishing these electronic switches for the interconnection makes this earlier developed system to be at high rate. This is one of the reasons why the vision of nationwide interconnection remained stymied before the advent of interconnectivity.

This situation demanded another approach: one that could simplify, automate and reduce the escalating costs involved in maintaining many-to-many connections. This prompted the government to issue an independent license to a clearinghouse that would encourage unfettered competition. Small operators can now communicate with bigger ones especially the mobile operators with a strict regulation that all must abide with.

The emergence of workable approach to solve the problem of nationwide and global interconnection application service provider (ASP) serves as a centralized, automated clearinghouse for carrier communications and transactions. This ASP receives all messages, and order and preorder information, automatically translates them to the right protocols, and directs them to the appropriate carriers. The business rules of all participating network operators are programmed into the clearinghouse, where they are updated as soon as a carrier's issue changes. Service providers can access the ASP via traditional gateways or the Internet [4].

Interconnecting mobile and fixed line requires guidelines which govern the operation and the process that gives a clear arrangement for interconnection and provision of services among the licensees. The concept of Interconnect Exchange cum Inter-Carrier Billing Clearing House which serves as a media Gate-way among the connected operators with the use of general acceptable Signaling protocols Signaling Switching Number 7 (SS7) an application which runs in any of the modern Telecommunication Switches. This has help in dealing with the dispute which might arise as a result of full competition in the telecommunications sectors in Nigeria [2].

Call Convergence Services are another type of services that involve licensed operators carrying information from its network (Fixed or Mobile) to another network operator either locally or internationally. It is used by individual operators

depending on the coverage of the license type like; (i) Public Switched Telephone Network (PSTN), (ii) Public Mobile Networks, and (iii) Global Mobile Satellite Services and Very Small Aperture Terminal (VSAT).

Currently, (a) Call Termination (b) Call Transit (c) Call Origination (including carrier selection and carrier pre-selection) and (d) Intelligent Network Calls (using Number Translation Services) are the different call convergence services applicable to the telecommunication sector. Here call termination can be defined as a service where an Operator receives voice band calls from another interconnected Operator, and then terminates (or completes) the calls within its own Public Telecommunications Network. Also a call transit service is defined as a service where an operator acting as a Clearing House receives voice band calls from one operator and routes them to the network of another connected Operators. It does not originate or terminate the call within its own network. Besides, call origination service is a service that provides carrier selection and carrier pre-selection is a service provided for a chosen user. Thus an operator makes calls to an interconnected operator but the originating operator does not bill the caller a retail tariff, but instead bills the other operator at an interconnect rate for originating the call. The call could be for any destination and will not necessarily terminate on the network of the operator who receives the call [5].

3 Traffic Analysis and Evaluation of System Interconnect.

The technical aspects which relate to the provision of physical facilities to enable networks connected to communicate with each other and transfer intelligence seamlessly across their boundaries. All the interconnected networks must come to term as regard technical issues to ensure inter-operability, particularly in technical matters.

There are basically two levels of interconnection:

- i. The horizontal level being the links between different telecommunication networks.
- ii. The vertical level being the links between ISPs. This vertical level links customers to the Internet through any of the telecommunication networks.

There is a standard deal with Interconnect issues which is related to the physical connection of separate networks to allow users of those networks to communicate with each other. This brought about the idea of inter-operability. Interoperability is the ability of two or more systems to work in concert with one another to achieve a predictable result. It is the ability of diverse systems made by different vendors to communicate with each other so that users do not have to make any necessary adjustments to account for differences in products or services [6].

All network and transmission equipments for interconnectivity must meet the minimum technical requirement. The Switching and Transmission systems must be designed on the internationally accepted Open Network Architecture (ONA) specifications and installed to the Open Network Provisioning standards. Alternate Routing / Diversity should be provided by either party in accordance with standard network engineering practices. This is just to improve the ASR and the SCR among the operators so as to have a good QoS.

Traffic is defined as the amount of data or the number of messages transmitted over a circuit

during a given period of time. Traffic also includes the relationship between call attempts on traffic-sensitive equipment and the speed with which the calls are completed. In this work, traffic analysis is used to ascertain the level of CCR and ASR in the interconnect system. A properly engineered network must have low blocking probability and high circuit utilization. There are many factors that should be taken into account when analyzing traffic. The most important factors are:

- Traffic load measurement
- Grade of service
- Traffic types
- Sampling methods. [7]

4 Traffic load measurement and data collection

The measured QoS parameters (data) as supplied by the ICH involve taking the readings for a period of one week at 15-minute intervals. These intervals are important because they are used to summarize the traffic intensity over a period of time. Taking the measurement throughout the day, we can find the peak hour traffic in any given day. The method used in taking the readings are ; (i) Daily Peak Period (DPP) which records the highest traffic volume measured during the day, (ii) Fixed Daily Measurement Interval (FDMI) which only requires measurements taken only during the predetermined peak period of the day when traffic pattern is predictable. Table 2 shows the QoS parameters collected from ICH which are recorded for ascertain the level of network performance of cellular networks interconnect in Nigeria. The traffic data collected was carries out for seven days of the week (Monday to Sunday). Table 1 is an extract from the entire data collected from the clearing house in August 2011.

Network	ASR Inc. (%)	ASR Out (%)	CCR (%)	CSS Inc.	CSS Busy	CSS Congest.
---------	--------------	-------------	---------	----------	----------	--------------

				dial		
Day One 15/08/2011 (Mon)						
BCETMN	33.85	35.46	100	1644	7568	8462
BETIMN	16.73	29.38	99.99	388	39448	1616
BGLOMN	33.62	14.17	100	902	22836	618696
BMTNAP	37.53	0	0	0	0	0
Day Two 16/08/2011 (Tue)						
BCETMN	33.44	35.37	99.99	857	3920	5519
BETIMN	10.7	29.57	100	156	19913	1257
BGLOMN	33.63	11.38	100	450	11105	352302
BMTNAP	37.33	0	0	0	0	0
Day Three 17/08/2011 (Wed)						
BCETMN	33.26	35.23	100	700	3694	4983
BETIMN	16.4	29.81	100	169	20324	1762
BGLOMN	32.58	12.48	100	503	11741	325973
BMTNAP	36.99	0	0	0	0	0
Day Four 18/08/2011 (Thur)						
BCETMN	33.92	30.84	100	1328	7416	17062
BETIMN	15.77	29.32	100	324	38336	2970

BGLOMN	33.43	12.90	100	898	24662	692352
BMTNAP	34.57	0	0	0	0	0
Day Five 19/08/2011 (Fri)						
BCETMN	34.86	35.55	100	695	3579	3094
METIMN	14.14	29.46	100	189	19920	1766
BGLOMN	33.22	12.11	100	507	12022	400531
BMTNAP	37.36	0	0	0	0	0
Day Six 20/08/2011 (Sat)						
BCETMN	34.52	34.23	100	703	3843	3483
BETIMN	9.4	29.33	100	219	19302	647
BGLOMN	31.92	12.30	100	470	11682	387258
BMTNAP	36.23	0	0	0	0	0
Day Seven 21/08/2011 (Sun)						
BCETMN	100	32.57	100	525	3645	2720
BETIMN	100	6.44	100	249	18870	648
BGLOMN	100	29.44	100	450	11173	295844
BMTNAP	0	33.43	0	0	0	0

Source: ICH Nigeria, August 2011.

The Quality of Service Audit of the various mobile operators network that was conducted by *technical research and standard (TRS)* of NCC to determine the performance of networks during the month of (March 2009). The two main groups of operators considered are: (A) GSM and (B) CDMA

The KPIs considered relevant to show the state of the networks are crucial to service providers who are concerned about maintaining reliable and stable networks in order to satisfy their subscribers. The indicators are:

- i. Traffic Channel Congestion(TCHC) (Without Handover): This is a performance indicator that shows the congestion of the voice channel.

$$TCHC = \frac{\text{total calls congested}}{\text{total calls attempted}} \times 100\% \quad (1)$$

- ii. Handover: This is the process in which control of a call is passed from one cell to another while the MS moves between calls.
- iii. Percentage of utilization (% TCH): This is the number of minutes of call that occurs at the

same time over the maximum traffic permitted with a grade of service (GOS) equal 2%.

$$\%TCHU = \frac{\text{total minutes of calls}}{\text{total period specified}} \times 100\% \quad (2)$$

- iv. Call Setup Success Rate (CSSR): Call Setup is an exchange of signaling information in the call process that leads to traffic channel seizure. The rate is the number of unblocked call attempts divided by the total number of call attempt

$$CSSR = \frac{\text{number of unblocked calls}}{\text{total number of calls attempted}} \times 100\% \quad (3)$$

- v. Call Drop Rate (CDR): The Call Drop is a premature termination of calls before being released normally by either the caller or called party. That is before the exchange of Release Message and Release Complete Message in signaling flow.

$$CDR = \frac{\text{unsuccessful calls}}{\text{total calls attempted}} \times 100\% \quad (4)$$

Following the monitoring activities carried out on the identified GSM operators mentioned, the performance statistics for the month of March 2009 are presented in Table 2. The figures represent network-wide average values of the Key Performance Indicators (KPIs) [8].

Table 2: Performance Statistics (QoS Indicator) of GSM Network Operators in Nigeria 2009

S/N	QoS Indicator	MTN (%)	GLO (%)	AIRTEL (%)	EMTS (%)
1	TCH Congestion (without	4.17	0.89	1.26	0.16

	Handover) Rate				
2	TCH Congestion (with handover) Rate	20.02	5.12	10.55	0.23
3	Call set-up Success Rate	81.28	94.53	87.18	98.09
4	Drop Call rate	1.42	1.63	0.91	1.12
5	SDDDH Congestion Rate	4.17	2.37	1.52	0.53

(Source: Consumer Affairs Bureau of NCC , 'March 2009')

5 Result of the Analysis

In today's competitive telecommunication environment, emphasis has shifted to delivery of exceptional services while maintaining a low cost operation. The telecommunication service providers are faced with the challenge to deliver high quality service, Innovative at a time of rapid change and cost pressure. Improved network availability and enhanced quality of service of network generated performance data can play significant roles in attaining the goal of interconnectivity.

Performance data is a resource for the information needed to manage the networks in a proactive way so that service problem can be addressed before it affects the customer. Much of raw data about various network elements is not useful until it is translated into real information.

Interconnection accounts for between 20%- 70% of operating costs, depending on the particular product. If regulators fail to understand, they can distort market entry signals, invalidate investments and encourage the abuse of dominant positions to the detriment of customers.

Interconnectivity is ideally settled among operators by themselves with the industry regulator, Nigerian Communications Commission (NCC) stepping in only when there is a serious dispute or crisis. The first result in figure 3 shows plot of the call completion ratio versus the different days for networks. This is the ratio of successfully completed calls to the total number of attempted calls at any given point of a network.

From the graph and analysis taken, Airtel had an average of about 75% of Call Completion Rate recorded for all the GSM operators within the network. Usually, for an effective network, the call completion rate should be 100%. The higher the percentage, the better.

Figures 4 and 5 show the graphical relationships between the answer seizure ratio with the number of days for different networks. While figure 4 depicts the different incoming answer seizure on different days, figure 5 shows different outgoing answer seizure ratio on different days.

Since busy signals and other rejections by the called number count as call failures, the calculated ASR value can vary depending on user behavior. Lower ASR may be caused by the Far End switch congestion, not Answering by called party and busy number at far end; all these factors bring a low (Poor) ASR which is uncontrollable by the operators.

The ASR is a measure of network quality defined in ITU SG2 Recommendation E.411: International network management - Operational guidance.

In Nigeria the communications commission (NCC) has set the licensing standards of Inward ASR requirements of 55% and outward ASR requirements of 66%. [7]

We are well aware that the battle for new subscribers has never been hotter. Increased competition, new digital technologies, enhanced

services, creative billing programs and increased churn rates have intensified the fight. Now more than ever we need to concentrate efforts on keeping your hard-won customers happy with their service.

Because subscribers, by and large, don't understand the technical jargon related to wireless services, they tend to judge the performance of your network at a very simple level- their ability to successfully place and end a call. Because of the way wireless services are marketed, customers have come to expect no difference in the performance of a wireless service compared with that of fixed-line service. About the only problem they will tolerate when placing a call is a busy tone from the called party, and even then not all the time.

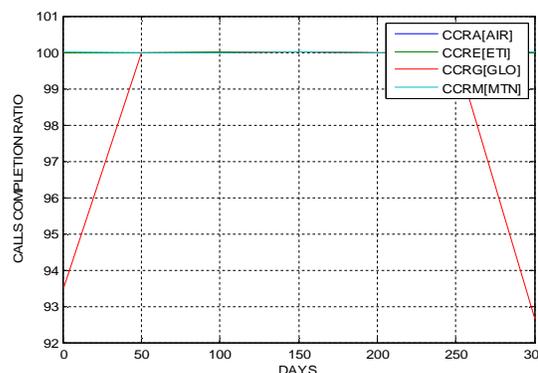


Figure 3: Graph of different Call Completion Ratio on different days

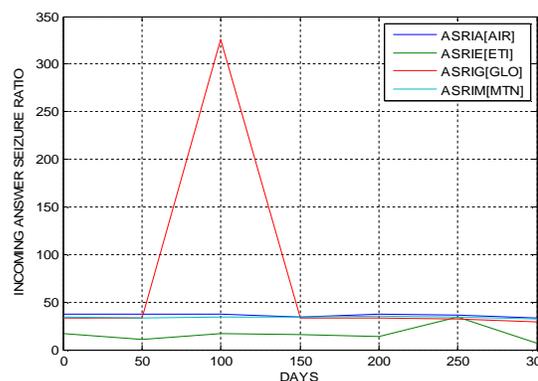


Figure 4: Graph of different incoming Answer Seizure Ratio on different days

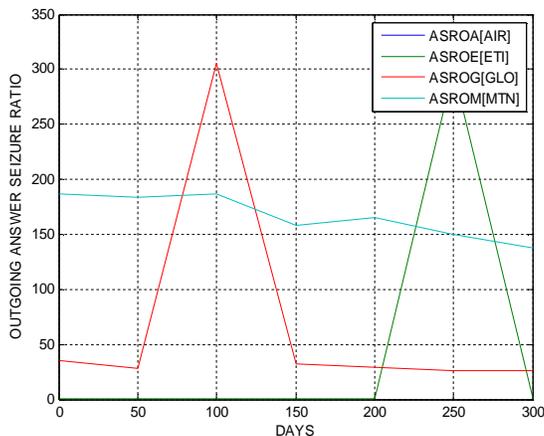


Figure 5: Graph of different outgoing Answer Seizure Ratio on different days

Unfortunately, reality does not always match expectations when it comes to wireless, which results in customers complaining about poor performance of the service. They eventually will churn if the perceived bad performance continues. It is the job of the communication engineers to see that the customers are able to place calls satisfactorily and hang up in their own time, regardless of the technical reasons for call-completion problems.

A range of factors can result in a customer's failing to complete a call satisfactorily. Although the percentages below will vary from network to network, they are a fair indication of the breakdown of call-completion problems facing networks and their subscribers:

- Handoff problems/dropped calls (29%): Incorrect neighbor lists, incorrect network handoff parameters, settings or hysteresis resulting in dropped calls
- Interference/quality (25%): Co-channel or adjacent channel or from external sources
- Lack of capacity/blocking (22%): Lack of capacity channels, trunks or resources,

particularly at peak traffic times, blocking access to the network

- Absence of RF (15%): Caused by poor coverage, shadowing, no in-building coverage or network equipment failure
- Phone or subscriber problems (9%): Poorly performing subscriber units, damaged or out of specification; poorly educated customers using handsets incorrectly by not raising antennas, choosing ill-advised locations to attempt calls or trying to place calls while outside advertised network coverage boundaries.

It is good network management to know in advance from the customers the problems they may find when using their various network. The network operators have the ability to check quickly on the reported problems. These network problems can be monitored via regular network audits or technical reviews.

6 Conclusion

From the results of this study, it is evident that the four GSM operators in Nigeria are not too far from providing reliable services. None of the GSM operators considered in the study has up to 100% call completion success rate. This is an indication that the service retainability of all GSM networks in the country is not satisfactorily very high. In fact, less than 90% of subscribers in each network get their calls completed before the call drop. This is an indication that the call drop rates on the networks are somehow low. The service retainability shows how long a subscriber stays on a network after the call has been set up or established; it is the measure of probability that an established call will not disconnect while conversation is in progress.

Likewise, the network accessibility on all the networks considered is low with half of the subscribers in each of the network need to dial three or more times before having access into various networks. This is an indication that the

congestion rates on the networks are high. It shows that all the GSM operators considered have more subscribers but lack sufficient equipment to support their daily increasing customer base. In addition, the result of the study as shown in the figures 3, 4, 5 shows that the bandwidth offered to the operators for network coverage for the entire nation is however overstretched. The study also reveals better performance of all the networks in terms of service integrity most especially in voice quality.

From the available data and the result of the analysis, it can be concluded that the QoS and overall performance of the GSM operation in Nigeria unsatisfactory and fairly reliable. It is an indication that Nigerians are yet to really enjoy fully the impact of GSM as a new effective means of telecommunication. In order to correct this ugly situation in the country and any other countries with similar situation, we are concluding this work by advising the network operators to optimize the allotted spectrum efficiently by building more BTs. This will ensure increased capacity, though the interconnectivity will be degraded as a result of congestion.

REFERENCES

- [1] Nigerian Communications Commission (NCC), "*Interconnectivity among network providers in Nigeria, NCC quarterly appraisal*", 2005.
- [2] Interconnect Clearing House Nig. Ltd 'www. Interconnectnigeria.com', 2009
- [3] Rakesh Mittal, "*Essential guide to telecommunications*", India 2004.
- [4] Web ProForum, "*Tutorials on Telecommunications Interconnections*",

'www.iec.org', International Engineering Consortium, 2010.

- [5] Nigeria Communications Commission, "*Interconnection Guideline*", October 14th, 2002.
- [6] Nigerian Communications Commission (NCC), "*Technical standard for Interconnectivity of Network*", NCC 2005
- [7] Aluko Mobolaji E., "*(Inter)Connectivity Issues in Nigeria: Some Perspectives*", Nigeria Telecomm summit, London Sept. 2004.
- [8] Nigeria Communications Commission (NCC), "*Performance Report of Mobile Operators Network (GSM and CDMA) in Nigeria for the Month of March 2009*", NCC Publications 2009.

Authors' Profiles

- **Ohaneme Cletus O.** is a lecturer in the department of Electronic and Computer Engineering of Nnamdi Azikiwe University, Awka, Nigeria. He is currently pursuing his PhD degree in Communication Engineering in the department of Electrical/Electronic Engineering of Enugu State University of Science & Technology Enugu, Nigeria.
E-mail: engrohaneme@yahoo.com
- **Akpado Kenneth A.** holds PhD in Computer and Control Engineering from Nnamdi Azikiwe University Awka, Nigeria. He is currently a lecturer in the department of Electronic and Computer Engineering, Nnamdi Azikiwe University, Awka, Nigeria.
E-mail: kenakpado@yahoo.com
- **Alumona Theophilus L.** is a lecturer in the department of Electronic and Computer Engineering, Nnamdi Azikiwe University Awka, Nigeria. He is currently pursuing PhD degree in Communication Engineering in the same department and University.
E-mail: hontheo@yahoo.com
- Emenike Stephanie I. holds B.Eng. in Electronic and Computer Engineering

(Communications option) from the
department of Electronic and Computer
Engineering, Nnamdi Azikiwe University,
Awka, Nigeria.
E-mail: ije_emenike@yahoo.com .

A Study of Clinostomum affecting *Oreochromis niloticus* in small water bodies in Eldoret-Kenya

Ochieng V. O., Matolla G.K., Khyria S.K.

Abstract - Two hundred and seventy nine specimens of *Oreochromis niloticus* from Kerita and Kesses dam were examined for Clinostomum parasites. Four hundred and thirty eight parasites were found on the two hundred and seventy nine fish specimens. This was done between December 2010 and February 2011.

The gut of *Oreochromis niloticus* were opened and checked for Clinostomum parasites, the number of the parasites found were recorded. The sex, total length and weight of fish were also recorded.

The parasites were mostly found attached to tissue behind the buccal cavity; this was associated to presence of adequate air supply (Coulbaly, 1995). This made it more probable that the Clinostomum parasite in this study was *Clinostomum tilapiae*.

The site of attachment on the fish tissue showed a cyst, the metacercariae is suspected to make this cyst as a form of protective mechanism to wall off and prevent displacement. The cyst was also concluded to have been produced by the fish as a defense mechanism to prevent further damage of tissue by the parasite

Kesses and Kerita dam had a parasitic prevalence of 75.71% 59.14% respectively; this difference was associated with the different human activities existing between the two dams, causing water quality differences.

In both dams, Male fish showed high intensity than female fish, this was associated with the breeding habits of *O. niloticus* which has the male fish spending long time in the shallow waters

Index terms: Clinostomum, *Oreochromis niloticus*, prevalence, Mean intensity, condition factor.

-
- Ochieng O. Vincent is currently pursuing masters degree program in Aquaculture, PH-+254729645504. E-mail: ochieng.vincent@ymail.com
 - Matolla G. K. is currently a PhD student and a lecturer at Chepkoilel University College.
 - Khyriah S. K. is currently pursuing masters degree program in fisheries management.
- (All authors are currently in Chepkoilel University College)

1 INTRODUCTION

FISH parasites are important to fish farming since the effects associated with inherent losses can be overwhelming. The environmental modifications and biological factors associated with fish farming in most cases encourage the development of fish pathogens. This has led to the observed cases of fish pathogens being more in farmed conditions unlike in natural ecosystems (Needham, 1978)

It has also been observed a great range of zoonotics associated with fish parasites, which can be managed by an understanding of the parasite diversity in a region. Ichthyozoonoses are other factors that necessitates studies of this kind (Dally, 1991)

The Digenea parasites are the main endo-parasites of teleost fishes; the greater majority of these fish are susceptible to infections with different stages of trematodes parasites (Masengo, 1989). The nature that has led to the poor understanding of the digenea parasites, including their diversity, ecology and biology include; the incomplete collection of trematodes of fishes, the standard of the taxonomy is highly variable and generally poor, there is a grave lack of underlying life-cycle information for several families and most species, and there have been relatively little work done to investigate these processes (Daly & Singleton, 1994).

There is no general pattern in the evolution of the digeneans (Moller, 1986); some genera appear closely co-evolved whereas others show no discernible fidelity to host taxa. Further, closely co-evolved genera may occur within the same family as others, with only ecological specificity to their hosts. (Needham, 1978).

The essence of this study is therefore to provide a basic framework for future studies and research on the Clinostomum parasites affecting fish in Uasin Gishu County. This was done by enabling a clear illustration and comparison of the parasitic indices

between two dams and relating them with the fish health.

2 MATERIALS AND METHODS

2.1 Study Site

The study was carried out in Kesses and Kerita dam in Uasin Gishu district. Laboratory analyses were done in Chepkoilel University College. The dams are located along Eldoret-Nairobi road and the University is situated along Ziwa road.

Important to note is that the conditions around Kerita dam are similar to the ones in Kesses dam. Therefore, reference is only made to Kesses dam.

2.1.1 Location: Kesses and Kerita dam are located in Uasin Gishu district. Uasin Gishu District (Figure 1) shares common borders with other districts like Trans-Nzoia to the north, Marakwet and Keiyo to the east, Koibatek to the south-east, Kericho to the south, Nandi to the west and Lugari to the north-west. Uasin Gishu district extends between longitude 34o 50' and 35o 37' east and 0o 03' and 0o 55' north.

Kesses Dam lies at an altitude of 2,750 m. a. s. l. and is located 31 km south of Eldoret town and 6 km east of Moi University Main Campus on Cheptiret – Lessos road and 9 km south of the main Nakuru-Eldoret road. Kesses Dam is thus about 43 km from Chepkoilel University College.

2.1.2 Hydrology and size: Kesses Dam receives most of its waters through Rivers Tarakwa and Nderugut and both rivers enter the east of the dam through a swamp of Typha latifolia and Cyperus sp. The dam's only outlet is the Sambul River in the west. The dam has surface area of about 189 hectares and a volume capacity estimated at about 274 m³. It has an average depth of 3 meters. The catchment area is approximately 1,720 hectares and extends as far as northern Tinderet forest and Nabkoi forest.

2.1.3 Fish communities: Tilapiines were the first fishes to be introduced in Kesses Dam with further

addition of about 5,000 fingerlings of *Oreochromis niloticus* (Linnaeus 1758) in 1990 and 1996 and that *Barbus* sp were available in the dam. *Gambusia* sp was also encountered in Kesses Dam (Ochoki, personal communication).

Macrophytes community: The macrophyte community includes *Phragmites* sp, *Nymphaea indica*, *Cyperus papyrus*, *Potamogeton* sp, *Utricularia foliosa*, *Ceratophyllum* sp and *Najas* sp. The shallow parts of the dam are covered with dense submerged *Ceratophyllum demersum*. These in most cases cause obstruction to fishing activities, especially gill netting. The most noticeable aquatic macrophytes in the dam include *Typha latifolia*, *Cyperus* sp, *Elodia canadensis* and *Potamogeton* sp, which formed a large swamp in the eastern part of the dam and fringe swamps around the remaining parts of the dam. This was similar with Kerita dam although the two dams are quite some distance apart.

2.1.4 Land use: The main human activities in the drainage basin of Kesses and Kerita dam includes majorly subsistence-crop farming, horticulture, agro-forestry, forestry and livestock rearing.

2.2 Fish Sample

During the period December 2010 and February 2011, a total of two hundred and seventy nine fish were sampled from both dams with 186 from Kesses and 93 from Kerita dam for the whole study. The samples were obtained by random sampling of hosts by use of gill net. The gill nets were set overnight and fish removed from them the following morning.

2.3 Parasites

Evisceration was done in the field to avoid the transportation cost that comes with transporting live fish, and it was also reliable working with freshly killed fish to eliminate the possibility of decomposition.

Sex, total length and weight of each fish were recorded; this was to enable calculations of condition factor.

Fish was opened by a longitudinal incision along the belly by use of a sharp scissors. The internal cavity of the fish was carefully examined for digenean parasites by use of magnifying lens. The mouth cavity and buccal cavity were also observed.

2.4 Data Analysis

Statistical analysis using Minitab was then done to establish the relationship between parasitic preference and fish sex, difference between the prevalence on the two dams and the relationship between fish condition factor and parasitic mean intensity. Chi square was used to ascertain significance in the tested relationships.

2.5 Determination of Parasite Indices

This was calculated as per the formulae listed below (Bush *et al*, 1997).

Prevalence: This is the number of hosts infected with one of more clinostomum parasite. It is expressed as a percentage and gives the level of population that is infected.

$$\text{prevalence} = \frac{\text{number infected hosts}}{\text{total host population}} \times 100\%$$

Mean Intensity: This is the average intensity of individuals of infected hosts.

$$\text{Mean Intensity} = \frac{\text{Total number of clinostomum parasites in a given host}}{\text{The number of hosts infected}}$$

$$\text{Mean abundance} = \frac{\text{Total number of parasites in a single host}}{\text{Number of examined hosts (both infected /not)}}$$

3 RESULTS

There were observed various differences in the distribution of parasites in the two dams, this is was evident on sex preference as shown in figures 3-6. These differences are attributed to environmental factors and others to the biology of the varied sex as therein explained.

The results also showed that Kesses dam had higher parasitic prevalence than Kerita dam; this could be associated with the low numbers of fish in the dam and the less human activities impacting the area as compared to Kesses dam. This is illustrated in figure 2.

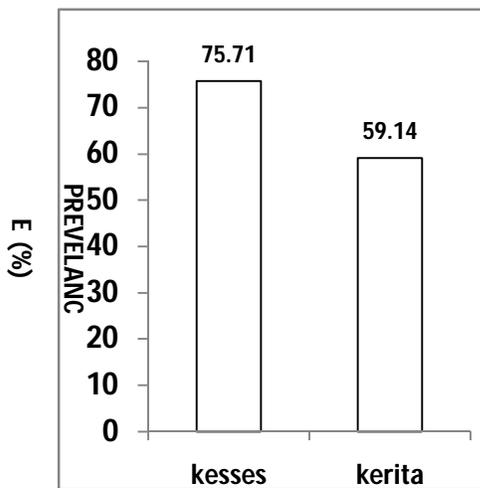


Figure 2: prevalence of infection of tilapia (*O. niloticus*) in Kesses and Kerita Dam by Clinostomum

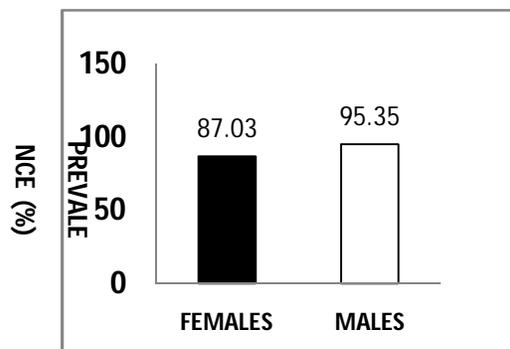
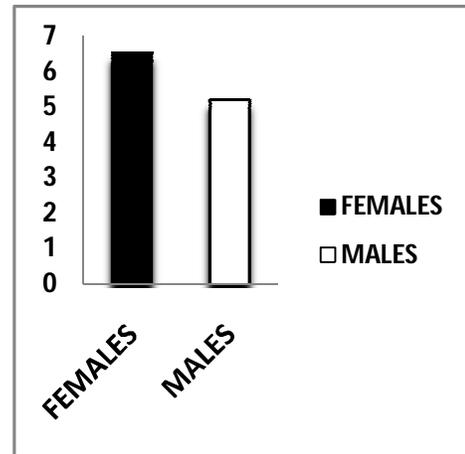


Figure 7: Mean intensity of Clinostomum parasites in males and female fish in Kerita Dam.



The graph bellow shows the preference by indicating the most and least affected sex.

Figure 3: Distribution of Clinostomum parasite in male and female in Kesses dam.

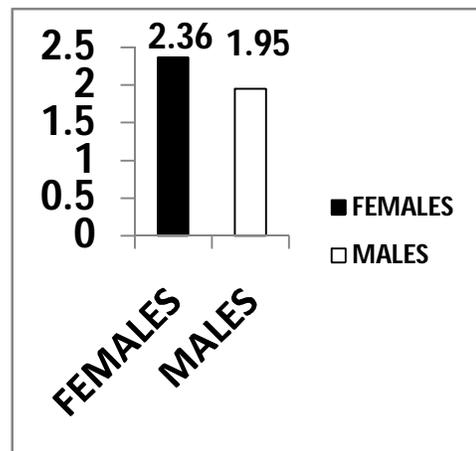


Figure 4: Mean intensity of by sex to show sex preference in Kesses dam results.

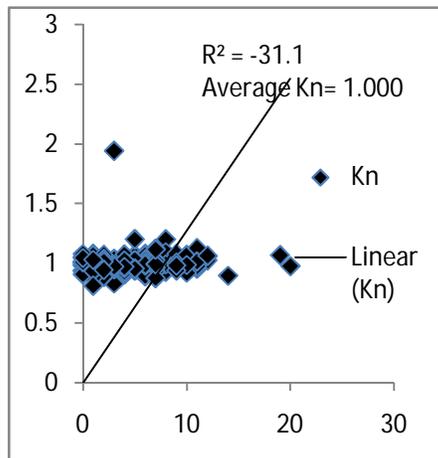


Figure 5: Relationship between condition factor and the number of parasites.

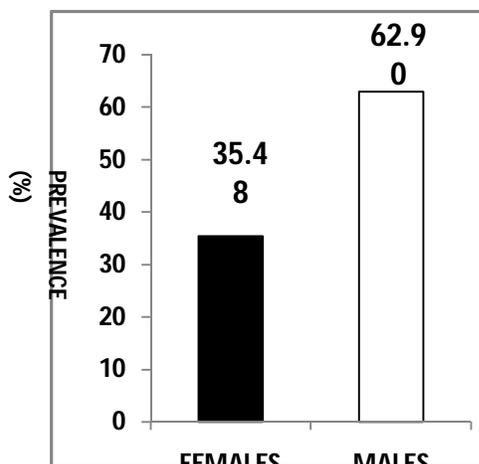


Figure 6: Prevalence of Clinostomum parasites in males and female fish in Kerita dam.

4 DISCUSSIONS

The study found out that the parasitic prevalence was higher in Kesses dam, 75.71%, than in Kerita dam, 59.14%. This meant that a greater proportion of fish in Kesses dam were more affected by the *Clinostomum* parasite.

This difference is highly associated with the presence of an urban center and School both releasing waste waters in Kesses dam. Kerita

dam is quite off the effects of urban life and therefore the water quality expected to be of fewer disturbances.

The parasitic intensity in Kesses dam was approximately 3 while that of Kerita dam was 1. This could also be associated with the water quality difference due to the influence of the urban center between the two dams and the presence of fewer aquatic birds observed in Kerita dam.

Clinostomids infection can have several effects on fish; among them being reduced growth, offset weight loss and prominent exophthalmia (Garacia *et al*, 1993; Yamaguti, 1933) however the results obtained in this study showed that the condition factor was not affected in any way by the parasites since there was observed a Pearson correlation value of 0.080 which represents no significant correlation between the number of parasites and the fish's condition. Equivalently, there was an average condition factor of 1, which is a reliable figure for well being.

Sex preference was observed in both dams with the male *O. niloticus* being favored by the *Clinostomum* parasite. This phenomenon is however still unclear but can be associated with the breeding habit of *O. niloticus*.

During breeding, male fish takes care of a territory around the nest, keeping off other fish (females not spawning included). During this period the males spend more time than the females in the shallow waters where the snails harboring the cercaria of *Clinostomum* are mostly found, and this is probably why the males had higher prevalence.

In this study the cysts were obtained behind the baccal cavity near the optic nerves. Ractiliffe, (1968) found that *C. tilapia* in reality has the physiologically adapted to turn upwards towards the mouth instead of going into the intestine because of the oxygen requirement. This conclusion by Ractiliffe therefore leads to an assumption that the *Clinostomum*

observed in the two dams were *Clinostomum tilapia*.

The observations also lead to a conclusion that the specie in question is *Clinostomum niloticus*; this is as provided by Ractiliffe, (1968).

The parasite on the fish may have a range of effects to the fish, and may stimulate other related responses to avoid damage to tissues; this may vary from mild inflammations to extremely severe narcotizing wound which may be critical (Hunter & Hunter, 1934). The cyst formed in the fish, may die off if its life cycle is not completed by predation on the fish by birds, the final hosts (Adeyemo et al, 2007).

In this study, the body cavity of *Clinostomum* was found as cysts, on the *niloticus*, shown in figure 8. Trematodes normally injure their host by mechanical damage resulting from consuming host tissues (Bullock, 1978). The parasite's cyst are produced by the fish tissues in reaction to attack by the parasite, this may conversely result into irritation of the affected area and surplus mucus reaction at the spot of attachment (Ukoli, 1866). The skin produces the cyst in a bid to prevent further penetration of the parasite into its tissues.

The position which the parasite decides to attach is majorly determined by availability of resources it requires for optimum survival. Amlacher (1966) observed that adult trematodes may not attack host organs by implanting in the tissues but only affix to well-situated sites where all required nutrients may be acquired with ease.

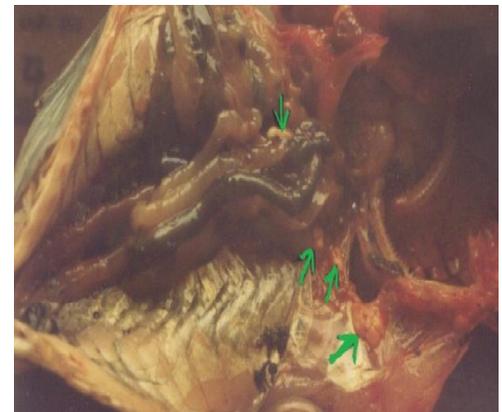


Figure 8: Body cavity of *O. niloticus* infested with *Clinostomum* parasites. Olsen, O.W (1985)

5 CONCLUSIONS

There was observed sex preference in the prevalence of the parasite. Generally, male fish had a higher prevalence than the female fish. This is associated with the male's behavior during breeding.

The parasitic indices for Kesses dam were generally higher than those in Kerita Dam, this was majorly associated with the environmental and human activities involved therein.

The parasitic abundance had no significant effect on the well being of fish, this was as represented by the 0.080 value of the Pearson correlation which a showed that there was no significance in correlation.

6 RECOMMENDATIONS

There need to be undertaken more studies on other existing parasites in the region to provide for more comparative study globally.

The study of *Clinostomum* parasites and other digeneans should be coupled with molecular biology to enable a more accurate description of their diversity and occurrences. This should lead to management solutions to control of these and other parasites in ways practical to the region.

ACKNOWLEDGMENT

The authors wish to thank Chepkoilel University College for the provision of facilities required during the study.

References

- Adeyemo A.O1 and Agbede, S.A2 (2007): Histopathology of Tilapia tissues harbouring *Clinostomum tilapiae* parasites. African Journal of Biomedical Research, Vol. 11 (2008); 115 – 118
- Amlacher E. (1996): Textbook of Fish Diseases. Translated by Canroy D.A and Herman, R.L.T.F.H publication inc. 308pp
- Bullock, A.M. (1978): Laboratory Methods. In fish Pathology. Roberts, R.J. (Editor).Bailliere Tindall Publ.235-266p
- Bush, A.O., Lafferty, K.D., Lotz, J.M., and Shostak, A.W. 1997. Parasitology meets ecology on its own terms: Margolis et al. revisited. J. Parasitol. 83: 575–583.
- Coulibay, N.D. et al. (1995): Larval *Clinostomosis* In Cichlid Fish in River Mahjong.
- Daly, J. J., B. DeYoung, and T. Hostetler. (1991): Hyperinfestation of smallmouth (*Micropterus dolomieu*) by the trematode *Clinostomum marginatum* in stream black bass. Proc. Arkansas Acad. Sci. 45:123.
- Daly,J. J., and j. Singleton. (1994): Estimating populations of yellow grub (*Clinostomum marginatum*) inchannel catfish (*Ictalurus punctatus*). Proc. Soc. Exptl. Biol. Med. 206:454.
- Garacia, M.L.J; Osorio, S.O and Constantino, F. (1993): Prevalence of parasites and the histological lesions they produce in Tilapia from Amela Lakes. Tecoman Colima. Vet Mex.
- Hunter, W.S. & Hunter, G.W. III (1934): Themiracidium of *Clinostomum marginatum*. Rud. Journal of Parasitology, 20, 132. Lake Kompeinga, Burkina Faso. A threat to Fishing and public Health. Francophones Sante
- Mashego, S.N (1989). Nematodes Parasites of *Barbus* species in Lebowa and Venda,
- Moller, H. and Anders, K. (1986). Diseases and parasites of Marine fishes. Kiel: Moller
- Needham, T and Wooten, R. (1978) Parasitology of Telesot in Fish Pathology. Robber
- Olsen, O.W (1985): Animal parasites. Their biology and Life Cycles Burgess Public Co. Minnesota. 100-106pp.
- Ractiliffe, , R.J. and Sommerville, C. (1968) "Disease of tilapias In: The biology and culture of tilapias (eds) R.S.V. Pullin and R.H. Lowe-McConnell). ICLARM Conference proceedings, Manilla Phillippines 7: 247-263 14
- South Africa S. Afr. J. Wild Res. 9(1): 35-37
- Ukoli, F.M.A. (1966): *Clinostomum tilapiae* sp., and *C. phalacrocoracis* Proc. Soc. Exptl. Biol. Med.
- Yamaguti, S. (1933): Studies on the helminth fauna of Japan. Part 1. Trematodes of Birds, Reptiles and Mammals. Japanese Journal of Zoology, 5, 1–134.

Multiconverter Unified Power-Quality Conditioning System: MC-UPQC

T.Charan Singh , L.Kishore, T.Sripal Reddy

Abstract—This paper presents a new unified power-quality conditioning system (MC-UPQC), capable of simultaneous compensation for voltage and current in multibus/multifeeder systems. In this configuration, one shunt voltage-source converter (shunt VSC) and two or more series VSCs exist. The system can be applied to adjacent feeders to compensate for supply-voltage and load current imperfections on the main feeder and full compensation of supply voltage imperfections on the other feeders. In the proposed configuration, all converters are connected back to back on the dc side and share a common dc-link capacitor. Therefore, power can be transferred from one feeder to adjacent feeders to compensate for sag/swell and interruption. The performance of the MC-UPQC as well as the adopted control algorithm is illustrated by simulation. The results obtained in PSCAD/EMTDC on a two-bus/two-feeder system show the effectiveness of the proposed configuration.

Index Terms—Power quality (PQ), MATLAB/SIMULINK, unified power-quality conditioner (UPQC), voltage-source converter (VSC)

I. INTRODUCTION

WITH increasing applications of nonlinear and electronically switched devices in distribution systems and industries, power-quality (PQ) problems, such as harmonics, flicker, and imbalance have become serious concerns. In addition, lightning strikes on transmission lines, switching of capacitor banks, and various network faults can also cause PQ problems, such as transients, voltage sag/swell, and interruption. On the other hand, an increase of sensitive loads involving digital electronics and complex process controllers requires a pure sinusoidal supply voltage for proper load operation [1]. In order to meet PQ standard limits, it may be necessary to include some sort of compensation. Modern solutions can be found in the form of active rectification or active filtering [2]. A shunt active power filter is suitable for the suppression of negative load influence on the supply network, but if there are supply voltage imperfections, a series active power filter may be needed to provide full compensation [3]. In recent years, solutions based on flexible ac transmission systems (FACTS) have appeared. The application of FACTS concepts in distribution systems has resulted in a new generation of compensating devices. A unified power-quality conditioner (UPQC) [4] is the extension of the unified power-flow Manuscript received October 23, 2007; revised November 13, 2008. Current version published June 24, 2009. Paper no. TPWRD-00628-2007. H. R. Mohammadi and A. Yazdian Varjani are with the Electrical Engineering Department, Tarbiat Modares University, Tehran, Iran (e-mail: mohammadi@modares.ac.ir; yazdian@modares.ac.ir). H. Mokhtari is with the Electrical Engineering Department, Sharif University of Technology, Tehran, Iran (e-mail: mokhtari@sharif.ir) Digital Object Identifier 10.1109/TPWRD.2009.2016822 controller (UPFC) [5] concept at the distribution level. It consists of combined series and shunt converters for simultaneous compensation of voltage and current imperfections in a supply feeder [6]–[8]. Recently, multiconverter FACTS devices, such as an interline power-flow controller (IPFC) [9] and the generalized unified power-flow controller (GUPFC) [10] are introduced. The aim of these devices is to control the

power flow of multilines or a subnetwork rather than control the power flow of a single line by, for instance, a UPFC. When the power flows of two lines starting in one substation need to be controlled, an interline power flow controller (IPFC) can be used. An IPFC consists of two series VSCs whose dc capacitors are coupled. This allows active power to circulate between the VSCs. With this configuration, two lines can be controlled simultaneously to optimize the network utilization. The GUPFC combines three or more shunt and series converters. It extends the concept of voltage and power-flow control beyond what is achievable with the known two-converter UPFC. The simplest GUPFC consists of three converters—one connected in shunt and the other two in series with two transmission lines in a substation. The basic GUPFC can control total five power system quantities, such as a bus voltage and independent active and reactive power flows of two lines. The concept of GUPFC can be extended for more lines if necessary. The device may be installed in some central substations to manage power flows of multilines or a group of lines and provide voltage support as well. By using GUPFC devices, the transfer capability of transmission lines can be increased significantly. Furthermore, by using the multilines-management capability of the GUPFC, active power flow on lines cannot only be increased, but also be decreased with respect to operating and market transaction requirements. In general, the GUPFC can be used to increase the transfer capability and relieve congestions in a flexible way. This concept can be extended to design multiconverter configurations for PQ improvement in adjacent feeders. For example, the interline unified power-quality conditioner (IUPQC), which is the extension of the IPFC concept at the distribution level, has been proposed in [11]. The IUPQC consists of one series and one shunt converter. It is connected between two feeders to regulate the bus voltage of one of the feeders, while regulating the voltage across a sensitive load in the other feeder. In this configuration, the voltage regulation in one of the feeders is performed by the shunt-VSC. However, since the source impedance is very low, a high amount of current would be needed to boost the bus voltage in case of a

voltage sag/swell which is not feasible. It also has low dynamic performance because the dc-link capacitor voltage is not regulated.

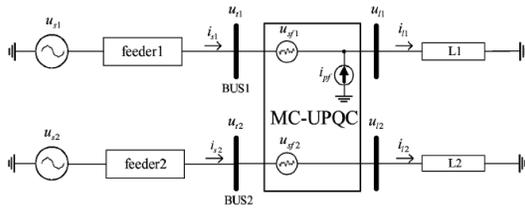


Fig. 1. Single-line diagram of a distribution system with an MC-UPQC.

In this paper, a new configuration of a UPQC called the multiconverter unified power-quality conditioner (MC-UPQC) is presented. The system is extended by adding a series-VSC in an adjacent feeder. The proposed topology can be used for simultaneous compensation of voltage and current imperfections in both feeders by sharing power compensation capabilities between two adjacent feeders which are not connected. The system is also capable of compensating for interruptions without the need for a battery storage system and consequently without storage capacity limitations.

II. PROPOSED MC-UPQC SYSTEM

A. Circuit Configuration

The single-line diagram of a distribution system with an MC-UPQC is shown in Fig. 1. As shown in this figure, two feeders connected to two different Substations supply the loads L1 and L2. The MC-UPQC is connected to two buses BUS1 and BUS2 with voltages u_{s1} and u_{s2} and respectively. The shunt part of the MC-UPQC is also connected to load L1 with a current i_{sp} . Supply voltages are denoted by u_{f1} and u_{f2} while load voltages are u_{l1} and u_{l2} . Finally, feeder currents are denoted by i_{f1} and i_{f2} and load currents are i_{l1} and i_{l2} . Bus voltages u_{s1} and u_{s2} are distorted and may be subjected to sag/swell. The load L1 is a nonlinear/sensitive load which needs a pure sinusoidal voltage for proper operation while its current is non sinusoidal and contains harmonics. The load L2 is a sensitive/critical load which needs a purely sinusoidal voltage and must be fully protected against distortion, sag/swell, and interruption. These types of loads primarily include production industries and critical service providers, such as medical centers, airports, or broadcasting centers where voltage interruption can result in severe economical losses or human damages.

B. MC-UPQC Structure

The internal structure of the MC-UPQC is shown in Fig. 2. It consists of three VSCs (VSC1, VSC2, and VSC3) which are connected back to back through a common dc-link capacitor. In the proposed configuration, VSC1 is connected in series with BUS1 and VSC2 is connected in parallel with load L1 at the end of Feeder1. VSC3 is connected in series with BUS2 at the Feeder2 end. Each of the three VSCs in Fig. 2 is realized by a three-phase converter with a commutation reactor and high-pass output filter as shown in Fig. 3. The commutation reactor and high-

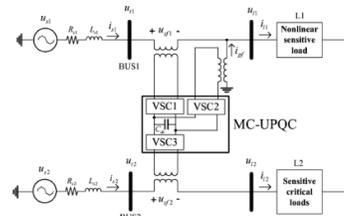


Fig. 2. Typical MC-UPQC used in a distribution system.

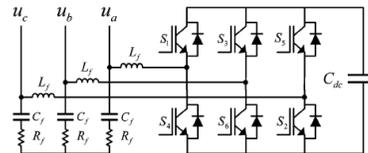


Fig.3.Schematic structure of a VSC.

pass output filter are connected to prevent the flow of switching harmonics into the power supply. As shown in Fig. 2, all converters are supplied from a common dc-link capacitor and connected to the distribution system through a transformer. Secondary (distribution) sides of the series-connected transformers are directly connected in series with BUS1 and BUS2, and the secondary (distribution) side of the shunt-connected transformer is connected in parallel with load L1. The aims of the MC-UPQC shown in Fig. 2 are: 1) to regulate the load voltage against sag/swell and disturbances in the system to protect the nonlinear/sensitive load L1; 2) to regulate the load voltage against sag/swell, interruption, and disturbances in the system to protect the sensitive critical load L2; 3) to compensate for the reactive and harmonic components of nonlinear load current. In order to achieve these goals, series VSCs (i.e., VSC1 and VSC3) operate as voltage controllers while the shunt VSC (i.e., VSC2) operates as a current controller.

C. Control Strategy

As shown in Fig. 2, the MC-UPQC consists of two series VSCs and one shunt VSC which are controlled independently. The switching control strategy for series VSCs and the shunt VSC are selected to be sinusoidal pulse width-modulation (SPWM) voltage control and hysteresis current control, respectively. Details of the control algorithm, which are based on the $\alpha\beta$ method [12], will be discussed later.

Shunt-VSC: Functions of the shunt-VSC are:

- 1) to compensate for the reactive component of load L1 current; Authorized licensed use limited to: Abdul Shaik. Downloaded on October 10, 2009 at 04:11 from IEEE Xplore. Restrictions apply. MOHAMMADI *et al.*: MULTICONVERTER UNIFIED PQ CONDITIONING SYSTEM 1681 Fig. 4. Control block diagram of the shunt VSC.
- 2) to compensate for the harmonic components of load L1 current;
- 3) to regulate the voltage of the common dc-link capacitor. Fig. 4 shows the control block diagram for the shunt VSC. The measured load current is

transformed into the synchronous reference frame by using (1) where the transformation matrix is shown in (2), at the bottom of the page. By this transform, the fundamental positive-sequence component, which is transformed into dc quantities in the d and q axes, can be easily extracted by low-pass filters (LPFs). Also, all harmonic components are transformed into ac quantities with a fundamental frequency shift (3) (4) where i_d and i_q are components of load current, i_{d0} and i_{q0} are dc components, and i_{d1} and i_{q1} are the ac components of i_d and i_q . If i_{d0} is the feeder current and i_{q0} is the shunt VSC current and knowing i_{d1} and i_{q1} , then i_{d0} and i_{q0} components of the shunt VSC reference current are defined as follows: (5) (6) Consequently, the d and q components of the feeder current are (7) (8) Fig. 5. Control block diagram of the series VSC. This means that there are no harmonic and reactive components in the feeder current. Switching losses cause the dc-link capacitor voltage to decrease. Other disturbances, such as the sudden variation of load, can also affect the dc link. In order to regulate the dc-link capacitor voltage, a proportional-integral (PI) controller is used as shown in Fig. 4. The input of the PI controller is the error between the actual capacitor voltage and its reference value. The output of the PI controller (i.e., v_{dc}) is added to the component of the shunt-VSC reference current to form a new reference current as follows: (9) as shown in Fig. 4, the reference current in (9) is then transformed back into the abc reference frame. By using PWM hysteresis current control, the output compensating currents in each phase are obtained

(10) *Series-VSC*: Functions of the series VSCs in each feeder are: 1) to mitigate voltage sag and swell; 2) to compensate for voltage distortions, such as harmonics;

3) To compensate for interruptions (in Feeder2 only). The control block diagram of each series VSC is shown in Fig. 5. The bus voltage is detected and then transformed into the synchronous reference frame using (11) where (12) (2)

Authorized licensed use limited to: Abdul Shaik. Downloaded on October 10, 2009 at 04:11 from IEEE Xplore. Restrictions apply.

1682 IEEE TRANSACTIONS ON POWER DELIVERY, VOL. 24, NO. 3, JULY 2009

and are fundamental frequency positive-, negative-, and zero-sequence components, respectively, and i_{d0} and i_{q0} is the harmonic component of the bus voltage. According to control objectives of the MC-UPQC, the load voltage should be kept sinusoidal with constant amplitude even if the bus voltage is disturbed. Therefore, the expected load voltage in the synchronous reference frame only has one value (13) where the load voltage in the abc reference frame is (14) The compensating reference voltage in the synchronous reference frame is defined as (15) This means in d and q should be maintained at while all other unwanted components must be eliminated. The compensating

reference voltage in d and q is then transformed back into the abc reference frame. By using an improved SPWM voltage control technique (sinePWMcontrol with minor loop feedback) [8], the output compensation voltage of the series VSC can be obtained.

III. POWER-RATING ANALYSIS OF THE MC-UPQC

The power rating of the MC-UPQC is an important factor in terms of cost. Before calculation of the power rating of each VSC in the MC UPQC structure, two models of a UPQC are analyzed and the best model which requires the minimum power rating is considered. All voltage and current phasors used in this section are phase quantities at the fundamental frequency. There are two models for a UPQC—quadrature compensation (UPQC-Q) and in phase compensation (UPQC-P). In the quadrature compensation scheme, the injected voltage by the series- VSC maintains a quadrature advance relationship with the supply current so that no real power is consumed by the series VSC at steady state. This is a significant advantage when UPQC mitigates sag conditions. The series VSC also shares the volt ampere reactive (VAR) of the load along with the shunt-VSC, reducing the power rating of the shunt-VSC. Fig. 6 shows the phasor diagram of this scheme under a typical load power factor condition with and without voltage sag. When the bus voltage is at the desired value, the series-injected voltage is zero [Fig. 6(a)]. The shunt

VSC injects the reactive component of load current, resulting in unity input-power factor. Furthermore, the shunt VSC compensates for not only the reactive component, but also the harmonic

components of the load current. For sag compensation in this model, the quadrature series voltage injection is needed as shown in Fig. 6(b). The shunt VSC injects in such a way that Fig. 6. Phasor diagram of quadrature compensation. (a) Without voltage sag. (b) With voltage sag. Fig. 7. Phasor diagram of in phase compensation (supply voltage sag). the active power requirement of the load is only drawn from the utility which results in a unity input-power factor. In an in phase compensation scheme, the injected voltage is in phase with the supply voltage when the supply is balanced. By virtue of in phase injection, series VSC will mitigate the voltage sag condition by minimum injected voltage. The phasor diagram of Fig. 7 explains the operation of this scheme in case of a voltage sag. A comparison between in phase (UPQC-P) and quadrature (UPQC-Q) models is made for different sag conditions and load power factors in [13]. It is shown that the power rating of the shunt-VSC in the UPQC-Q model is lower than that of the UPQC-P, and the power rating of the series-VSC in the UPQC-P model is lower than that of the UPQC-Q for a power factor of less than or equal to 0.9. Also, it is shown that the total power rating of UPQC-Q is lower than that of UPQC-P where the VAR demand of the load is high. As discussed in Section II, the power needed for interruption compensation in Feeder2 must be supplied through the shunt VSC in Feeder1 and the series VSC in Feeder2. This implies that power ratings of these VSCs are greater than that of the series one in Feeder1. If quadrature compensation in Feeder1 and in phase compensation in Feeder2 are selected, then the power rating of the shunt VSC and the series VSC (in Feeder2) will be reduced. This is an important criterion for practical applications. Based

on the aforementioned discussion, the power-rating calculation for the MC-UPQC is carried out on the basis of the linear load at the fundamental frequency. The parameters in Fig. 6 are corrected by adding suffix "1," indicating Feeder1, and the parameters in Fig. 7 are corrected by adding suffix "2," indicating Feeder2. As shown in Figs. 6 and 7, load voltages in both feeders are kept constant at regardless of bus voltages variation, and the load currents in both feeders are assumed to be constant at their rated values (i.e., and , respectively) (16)

Authorized licensed use limited to: Abdul Shaik. Downloaded on October 10, 2009 at 04:11 from IEEE Xplore. Restrictions apply. MOHAMMADI *et al.*: MULTICONVERTER UNIFIED PQ CONDITIONING SYSTEM 1683 (17) The load power factors in Feeder1 and Feeder2 are assumed to be and and the per-unit sags, which must be compensated in Feeder1 and Feeder2, are supposed to be and, respectively. If the MC-UPQC is lossless, the active power demand supplied by Feeder1 consists of two parts: 1) the active power demand of load in Feeder1; 2) the active power demand for sag and interruption compensation in Feeder2. Thus, Feeder1 current can be found as (18) (19) (20) (21) From Fig. 6, the voltage injected by the series VSC in Feeder1 can be written as in (22) and, thus, the power rating of this converter can be calculated as (22) (23) The shunt VSC current is divided into two parts.

1) The first part (i.e.,) compensates for the reactive component (and harmonic components) of Feeder1 current and can be calculated from Fig. 6 as (24) where is calculated in (21). This part of the shunt VSC current only exchanges reactive power (Q) with the system. 2) The second part provides the real power (P), which is needed for a sag or interruption compensation in Feeder2. Therefore, the power rating of the shunt VSC can be calculated as (25) where is calculated in (24). Finally, the power rating of the series-VSC in Feeder2 can be calculated by (26). For the worst-case scenario (i.e., interruption compensation), one must consider . Therefore (26) Fig. 8. BUS1 voltage, series compensating voltage, and load voltage in Feeder1.

IV. SIMULATION RESULTS

The proposed MC-UPQC and its control schemes have been tested through extensive case study simulations using MATLAB/SIMULINK. In this section, simulation results are presented, and the performance of the proposed MC-UPQC system is shown.

A. Distortion and Sag/Swell on the Bus Voltage

Let us consider that the power system in Fig. 2 consists of two three-phase three-wire 380(v) (rms, L-L), 50-Hz utilities. The BUS1 voltage contains the seventh-order harmonic with a value of 22%, and the BUS2 voltage contains the fifth order harmonic with a value of 35%. The BUS1 voltage contains 25% sag between s and 20% swell between s. The BUS2 voltage contains 35% sag between s and 30% swell between s. The nonlinear/sensitive load L1 is a three-phase rectifier load which supplies an RC load of 10 and 30 F. Finally, the critical load L2 contains a balanced RL load of 10 and 100 mH. The MC-UPQC

is switched on at 0.02 s. The BUS1 voltage, the corresponding compensation voltage injected by VSC1, and finally load L1 voltage are shown in Fig. 8. In all figures, only the phase waveform is shown for simplicity. Similarly, the BUS2 voltage, the corresponding compensation voltage injected by VSC3, and finally, the load L2 voltage are shown in Fig. 9. As shown in these figures, distorted voltages of BUS1 and BUS2 are satisfactorily compensated for across the loads L1 and L2 with very good dynamic response. The nonlinear load current, its corresponding compensation current injected by VSC2, compensated Feeder1 current, and, finally, the dc-link capacitor voltage are shown in Fig. 10. The distorted nonlinear load current is compensated very well, and Fig. 9. BUS2 voltage, series compensating voltage, and load voltage in Feeder2.

Fig. 10. Nonlinear load current, compensating current, Feeder1 current, and capacitor voltage. the total harmonic distortion (THD) of the feeder current is reduced from 28.5% to less than 5%. Also, the dc voltage regulation loop has functioned properly under all disturbances, such as sag/swell in both feeders.

B. Upstream Fault on Feeder2

When a fault occurs in Feeder2 (in any form of L-G, L-L-G, and L-L-L-G faults), the voltage across the sensitive/critical Fig. 11. Simulation results for an upstream fault on Feeder2: BUS2 voltage, compensating voltage, and loads L1 and L2 voltages. load L2 is involved in sag/swell or interruption. This voltage imperfection can be compensated for by VSC2. In this case, the power required by load L2 is supplied through VSC2 and VSC3. This implies that the power semiconductor switches of VSC2 and VSC3 must be rated such that total power transfer is possible. This may increase the cost of the device, but the benefit that may be obtained can offset the expense. In the proposed configuration, the sensitive/critical load on Feeder2 is fully protected against distortion, sag/swell, and interruption. Furthermore, the regulated voltage across the sensitive load on Feeder1 can supply several customers who are also protected against distortion, sag/swell, and momentary interruption. Therefore, the cost of the MC-UPQC must be balanced against the cost of interruption, based on reliability indices, such as the customer average interruption duration index (CAIDI) and customer average interruption frequency index (CAIFI). It is expected that the MC-UPQC cost can be recovered in a few years by charging higher tariffs for the protected lines. The performance of the MC-UPQC under a fault condition on Feeder2 is tested by applying a three-phase fault to ground on Feeder2 between s. Simulation results are shown in Fig. 11.

C. Load Change

To evaluate the system behavior during a load change, the nonlinear load L1 is doubled by reducing its resistance to half at 0.5 s. The other load, however, is kept unchanged. The system response is shown in Fig. 12. It can be seen that as load L1 changes, the load voltages and remain undisturbed, Authorized licensed use limited to: Abdul Shaik. Downloaded on October 10, 2009 at 04:11 from IEEE Xplore. Restrictions apply. MOHAMMADI *et al.*: MULTICONVERTER UNIFIED PQ

CONDITIONING SYSTEM 1685 Fig. 12. Simulation results for load change: nonlinear load current, Feeder1 current, load L1 voltage, load L2 voltage, and dc-link capacitor voltage. the dc bus voltage is regulated, and the nonlinear load current is compensated.

D. Unbalance Voltage

The control strategies for shunt and series VSCs, which are introduced in Section II, are based on the – method. They are capable of compensating for the unbalanced source voltage and unbalanced load current. To evaluate the control system capability for unbalanced voltage compensation, a new simulation is performed. In this new simulation, the BUS2 voltage and the harmonic components of BUS1 voltage are similar to those given in Section IV. However, the fundamental component of the BUS1 voltage is an unbalanced three-phase voltage with an unbalance factor of 40%. This unbalance voltage is given by (27) the simulation results for the three-phase BUS1 voltage, series compensation voltage, and load voltage in feeder 1 are shown in Fig. 13. The simulation results show that the harmonic components and unbalance of BUS1 voltage are compensated Fig. 13. BUS1 voltage, series compensating voltage, and load voltage in Feeder1 under unbalanced source voltage. for by injecting the proper series voltage. In this figure, the load voltage is a three-phase sinusoidal balance voltage with regulated amplitude.

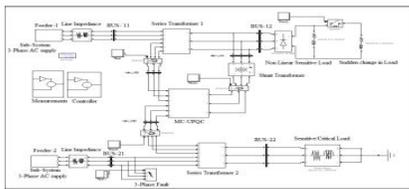


Fig4. Simulink model of distributed system without MCUPQC

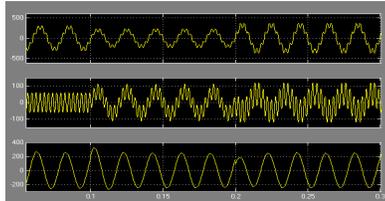


Fig4.2. BUS1 voltage, series compensating voltage, and load voltage in Feeder1.

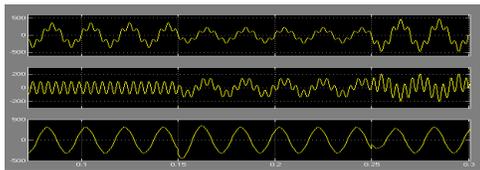


Fig4.3 BUS2 voltage, series compensating voltage, and load voltage in Feeder2.

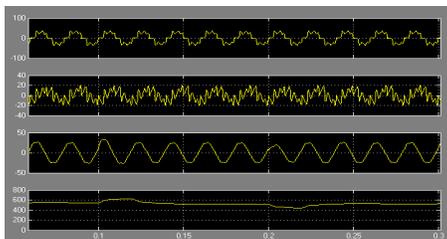


Fig. 4.4 Nonlinear load current, compensating current, Feeder1 current, and capacitor voltage.

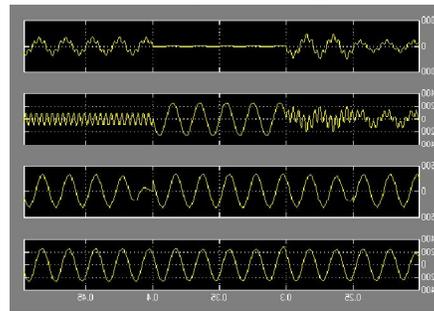


Fig. 4.5. Simulation results for an upstream fault on Feeder2: BUS2 voltage, compensating voltage, and loads L1 and L2 volt.

V. CONCLUSION

In this paper, a new configuration for simultaneous compensation of voltage and current in adjacent feeders has been proposed. The new configuration is named multiconverter unified power-quality conditioner (MC-UPQC). Compared to a conventional UPQC, the proposed topology is capable of fully protecting critical and sensitive loads against distortions, sags/swell, and interruption in two-feeder systems. The idea can be theoretically extended to multibus/multifeeder systems by adding more series VSCs. The performance of the MC-UPQC is evaluated under various disturbance conditions and it is shown that the proposed MC-UPQC offers the following advantages:

- 1) power transfer between two adjacent feeders for sag/swell and interruption compensation;
- 2) compensation for interruptions without the need for a battery storage system and, consequently, without storage capacity limitation;
- 3) Sharing power compensation capabilities between two adjacent feeders which are not connected.

REFERENCES

[1] D. D. Sabin and A. Sundaram, "Quality enhances reliability," *IEEE Spectr.*, vol. 33, no. 2, pp. 34–41, Feb. 1996.
 [2] M. Rastogi, R. Naik, and N. Mohan, "A comparative evaluation of harmonic reduction techniques in three-phase utility interface of power electronic loads," *IEEE Trans. Ind. Appl.*, vol. 30, no. 5, pp. 1149–1155, Sep./Oct. 1994.
 [3] F. Z. Peng, "Application issues of active power filters," *IEEE Ind. Appl. Mag.*, vol. 4, no. 5, pp. 21–30, Sep./Oct. 1998.
 [4] H. Akagi, "New trends in active filters for power conditioning," *IEEE Trans. Ind. Appl.*, vol. 32, no. 6, pp. 1312–1322, Nov./Dec. 1996.
 [5] L. Gyugyi, C. D. Schauder, S. L. Williams, T. R. Rietman, D. R. Torjerson, and A. Edris, "The unified power flow controller: A new approach to power transmission control," *IEEE Trans. Power Del.*, vol. 10, no. 2, pp. 1085–1097, Apr. 1995. Authorized licensed use limited to: Abdul Shaik. Downloaded on October 10, 2009 at 04:11 from IEEE Xplore. Restrictions apply. 1686 IEEE TRANSACTION POWER DELIVERY, VOL. 24, NO. 3, JULY 20[6H.

Impact of Biotic Interferences on Yousmarg Forest Ecosystem, Kashmir

BHAT MOHD SKINDER, ASHOK K. PANDIT

ABSTRACT— The present investigation of forest ecosystem was undertaken during May 2010 to Dec 2010. The main aim of the study was to analyze the impact of biotic interferences on Yousmarg forest ecosystem. The cumulative effect of the biotic interferences was significantly seen in the reduction of vegetation cover. The present study revealed that the prominent factor for the exploitation of the vegetation cover of the study area at herbaceous level is simply the overgrazing. The increasing disturbances not only disturb the plant species diversity, richness and evenness significantly but various plant species have been got completely eliminated from grazing area by different kinds of interferences like overgrazing, deforestation, forest fire, tourist impacts etc. It has been observed that about 10 species were absent in the grazing area where as all species were present in the fenced (protected) area .Most of the species have medicinal value like *Achillea millefolium*, *Fragaria nubicola*, *Taraxacum officinale*, *Prunella vulgaris*, *Potentilla reptans*, *Rumex hastatus*, etc. The data showed that the importance value index (IVI) of *Pinus wallichiana* at all the three sites like Nabrepathar, Mashid Nar and Chota Yous of Yousmarg area, gets constantly decrease from protected site to unprotected site while as there is over all decrease in the IVI of other two species like *Picea smithana* (spruce) and *Abies pindrow* (fir).

KEY WORDS- Yousmarg forest, ecosystem, biotic interferences, vegetation analysis, medicinal plants, vegetation cover and exploitation



1 INTRODUCTION

The vast and varied natural resources of the earth are essential for the sustenance and well being of mankind. Among the natural resources, the forest resources occupy a unique positioning as much as they play an important role in the development of society by providing a wide range of goods and services. Besides the economic amenities, the forests have diverse ecological roles as the forests are the primary source to regenerate productivity of land through recycling of nutrients, which may change the physicochemical features of soil favourable for plant growth [1]. Despite their economical and ecological importance, forests are facing multitude of impacts which have not only reduced their overall cover but have also lead to the extinction of some important species of both plants and animals. The degradation of forests is chiefly due to various anthropogenic pressures [2]. Overgrazing reduces plant leaf areas, which reduces interception of sunlight and plant growth. Plants become weakened and have reduced root length, and the pasture sod weakens. The reduced root length makes the plants more susceptible to death during dry weather. The weakened sod allows weed seeds to germinate and grow. If the weeds are unpalatable or poisonous, major problems can result [3].

Bhat Mohd Skinder is currently a M.Phil scholar (2012-2013) in P.G. Department of Environmental Science/Centre of Research for Development in University of Kashmir-190006, India (Jammu and Kashmir), PH- +91 9796371486. E-mail: mskbhat@gmail.com

Ashok K. Pandit – Head, P.G. Department of Environmental Science / and Director, Centre of Research for Development, University of Kashmir-190006, India (Jammu and Kashmir) PH- +91 9419091424. E-mail: prof_akpandit@yahoo.co.in

2 STUDY AREA-YOUSMARG



Yousmarg is an emerging destination for tourist spot, which is completely raw, pristine and still unspoiled forest ecosystem lying in the Budgam district of Kashmir valley, India. It lies between the geographical co-ordinates: 33°49'42"N, and 74° 39'59"E and at an elevation of 2712 m (a.s.l.). Yousmarg is about 30 kms from Budgam and about 47 km from Srinagar. Yousmarg is bandied by lush green grasslands, rivers and the backdrop of snow capped mountains could leave the travellers spellbound for a few seconds Yousmarg mesmerizes tourists with its scenic meadows, a sparkling reservoir and mountains comparable to European Alps. It has some of the highest peaks in Pir-Panjal range like Tatakoti 4725m, Romesh Thong 5000m and Sunset Peak 4746 m. The mighty river Doodh Ganga rises from these peaks and distributaries of the same flows into the reservoir (Named Sardemouj reservoir) the mighty river Doodhganga makes this destination more thrilling.

Nabrepathar site. This site is just half kilometer away from the Yousmarg main bowl and is to the west side of the main bowl

and lies between geographical coordinates. N 33° 50' 00.8'' E 074° 40' 03.4 and at an altitude 2439m (a.s.l)

Mashid Nar site. This site is also half kilometer away from the Yousmarg main bowl and is to the East side of the main bowl and lies between geographical coordinates of N 33° 49' 25.9'' E 074° 40' 10.3''

Chota Yous site. This site is one kilometer away from the Yousmarg main bowl and nearer to the settlement village Nagabal, south side and lies between geographical coordinates of N 33° 48', 33.7'' E 074° 41', 13.3'' and at an altitude 2309m(a.s.l)

3 METHODS

Survey data was collected from May 2010 - Dec 2010. The findings in this study are related to various aspects like human habitation, grazing pressure, erosion, deforestation etc. in Yousmarg forest areas were simply based on:

Vegetation analysis or community structure is an important aspect of ecology and enables us to understand as to how a community formed by organisms control the organism and influence their behavior and life history [4]. A community is a component as well as factor of environment. The vegetal analysis provides information regarding the interaction among species in a particular community as well as about the organization of the species within the community. Vegetation analysis was conducted during May 2010-Dec.2011 for all the three layers of the forest i.e. trees, shrubs and herbs. The species were identified by the Taxonomy Department of Kashmir University. The tree layer was analyzed by sampling 10 quadrates of 10 x 10 m size at each site. The size and

number of samples were determined using the method of [4],[5]. The abundance, density and frequency were calculated for the species. IVI was determined as the sum of the relative frequency, relative density and relative dominance for tree layer and Herb layer only. The shrub and herbs layers were analyzed by randomly placing 20 quadrates of 5 x 5m size and 1 x 1m size respectively at each site during the summer season. The diversity index at each site was computed by using Shannon- Wiener information function [6] and concentration of dominance by Simpson's index [7], evenness and richness index following [8],[9] respectively. Similarity between the grazing site and Protected (Fenced) site for Herbs was determined using Sorenson's index of similarity [10].

4 RESULTS

4.1 Phytosociological analysis

During the present study some 29 species of plants were collected among which 24 were identified as herbs, 2 as shrubs and 3 as trees (Table 1)

4.2 Herbs

In the first set of vegetation analysis frequency, density, abundance and IVI was compared between the protected and unprotected plots at each site so, to determine the intensity of impact of biotic factors. One site was selected where extensive grazing taking place. A total number of twenty random quadrates of 1x1m [4], were laid for herb analysis at grazing plot and fenced plot (protected area)

A total of 24 species were identified by taxonomy department of Kashmir University, among them only 14 species were present in grazing area and all the species were present in the fenced area. A perusal of data shows that in the Yousmarg forest division, the total frequency (1980), density (248.72) and abundance (271.87) are more at fenced area than the degraded or grazed plot (968, 162.5 and 191.5) (Figure 1 and 2). The importance value index (IVI) at the grazing area ranges from 72.19 for *Cyanodon* sp. to 6.54 for *Medicago* sp. While as it ranges from 40.39 again for *Cyanodon* sp to 3.98 for *Ranunculus* sp. at Fenced area (protected plot), (Figure 1 and 2).

The species diversity is more at fenced area (2.97) as compared to grazing area (2.33), similarly species richness (5.75) and evenness index (2.15) are in higher values. So far as deforestation is concerned the values of frequency, density and abundance gets decrease from protected site to degraded area. The frequency range at the three protected Nabrepathar, Mashid Nar and Chota Yous is 75 to 100, 50 to 75 (Nabrepathar and Mashid Nar) and 25 to 100 (Chota Yous) at degraded areas. The density at protected site Nabrepathar ranges from 1.75 to 3.25, at site Mashid Nar, 1.25 to 2.75 and at site Chota Yous, it is 1.66 to 6.5, while as at degraded sites the values are 0.75 to 1.75, 0.75 to 1.25 and 0.25 to 2.75 in the respective sites. Similarly the values of abundance at protected sites are 2.00 to 3.25, 1.66 to 2.75 and 1.66 to 6.5. While at degraded areas these values range from 1.50 to 2.33, 1.00 to 1.66 and 1.00 to 2.75, in the respective sites (Table 3). The data showed that in the

as compared to the grazing areas 3.70 and 2.03 in respective indices. While as the dominance index at grazing area was 0.88 as it is highest 0.93 at the fenced (protected) site. So far as similarity index is concerned, Sorenson index shows only 57.42% similarity between the two sites (Table 2)

4.3 Shrubs

One site was selected (site 2) having a good coverage of shrub vegetation was the focus of present investigation while the other two sites had scanty shrub vegetation. Comparison of shrubs at protected site and degraded site on the basis of frequency, density and abundance shows a decline of values from degraded to protected site as given in the Figure 3.

4.4 Trees

Nabrepathar site of Yousmarg forest, the IVI for the protected site was 74.85 for *Pinus wallichiana*, 118.91 for *Picea smithana* and 74.85 for *Abies pindrow* as against the values of 69.47, 121.15 and 109.30 for the degraded areas in the respective spp. While as in the Mashid Nar site IVI in the protected site came to be 76.52 for *Pinus wallichiana*, 118.45 for *Picea smithana* and 110.99 for *Abies pindrow* as against the values 75.00, 104.08 and 120.00 in the degraded areas in the respective sp. Similar is the case with Chota Yous site, where IVI in the protected site was 152.41 for *Pinus wallichiana*, 94.38 for *Picea smithana* and 53.17 for *Abies pindrow* while as in degraded areas these values were 112.00, 153.00 in the respective sp. (Fig.4,5 and 6)

5 DISCUSSION

The present study revealed that the prominent factor for the exploitation of the vegetation cover of the Yousmarg forest study area at herbaceous level is simply the overgrazing. This not only become the cause for the destruction of herbaceous cover but also becomes the secondary biotic factor for the soil erosion, as sometimes grazers root out herbs and exposed the soil which ultimately gets eroded during the rainy season and the forest area generally located at some slope, thus slope becomes the catalyst in the process of soil erosion. Overgrazing can increase soil erosion, reduced soil depth, soil organic matter, and soil fertility, hurt the land's future productivity. Soil fertility can be corrected by applying the appropriate lime and fertilizers. However, the loss of soil depth and organic matter takes years to correct. Their loss is critical in determining the soil's water-holding capacity and how well pasture plants do during dry weather similar concepts were discussed by [3].

It has been observed that about 10 species were absent in the grazing area where as all species were present in the fenced (protected) area (Fig.1 and 2) Most of the species have medicinal value like *Achillea millefolium*, *Fragaria nubicola*, *Taraxacum officinale*, *Prunella vulgaris*, *Potentilla reptans*, *Rumex hastatus*, etc. The striking feature of the present study is that the total frequency, density and abundance gets decrease from fenced area to grazing area and some species especially medicinal species were totally absent in the grazing area (Fig. 1 and 2). This shows a serious concern of the overgrazing. However,

the values of IVI is not just maximum at fenced plot (protected) for some species but also in grazed plot (unprotected) (Figure 1 and 2), Similar observations have also been reported by other authors elsewhere [11], [12], [13]. The IVI of the *Cyanodon dactylon* (Fig. 1 and 2) should be higher in the fenced area as it is the dominant species not only at the study area but also in the Kashmir valley (common observation) but showed a reverse trend, the reason may be attributed to competition among various herbaceous species on a limited area on the account of nutrient availability, moisture content of soil, space and light in the fenced area. It has been observed that similarity Sorenson index shows only 57.42% similarity between the two sites. (Table 1) while as all the indices like dominance index, species diversity, species richness and evenness index are in higher values as compared to the grazing areas (Table 2). So the striking feature of this study at herbaceous level is that anthropogenic interference decreases the species richness or species diversity or evenness index or dominance index, similar observation was made by [12], [14], [15].

So far as deforestation is concerned the values of frequency, density and abundance gets decrease from protected site to degraded area (Table 3). Similar results were also obtained by [2], [16], [17], [18]. The present data shows some variation in the three tree species present in the Yousmarg area (Fig.4, 5 and 6). The concept of Importance Value Index has been used extensively as a means of assessing the biological contribution of species to the forest community [19]. The data showed that the importance value index (IVI) of *Pinus*

wallichiana at all the three sites like Nabrepathar, Mashid Nar and Chota Yous of Yousmarg area, gets constantly decrease from protected site to unprotected site while as there is overall decrease in the IVI of other two species like *Picea smithana* (spruce) and *Abies pindrow* (fir) (Fig.4, 5 and 6). According to the local people of the Yousmarg area *Pinus wallichiana* (local name "Kayur") is most vulnerable for cutting as it is good for construction, While as *Picea smithana* (local name "Kachul") and *Abies pindrow* (local name "Rayul") collective name of both species is "Budul" used for mainly fire wood. Illicit cutting and removal of timber has now become a prevalent practice, which impacts the regeneration potential of the study sites. The problem with these forms of forest disturbances is that the plants often do not have time to recover adequately and thus, these impacts affect the ecosystem succession of the communities [20]. The IVI also indicates that *Picea smithana* and *Abies species* is dominant at higher altitudes (2440m) both at protected as well as degraded areas but the dominant species at lower altitude (2309m) is *Pinus wallichiana* (blue-pine) at protected site (Fig.4,5 and 6). So the results revealed that two trends are being observed at the Yousmarg forest area. At higher altitudes *Picea* species and *Abies* species are dominant while as *Pinus* species is dominant at lower altitudes. Hence, the

CONCLUSION

The study reveals that the grazing pressure is prime factor for the vegetation loss at herbaceous level. The most prospective loss is in the form of medicinal plants .The tree

results obtained in the present were found to be in consonance with the results obtained by [21].

The density and abundance of shrubs at degraded area are more than the protected areas (Fig.3). The most striking features were an abrupt decline in the number of primary forest regenerative species, especially *Pinus wallichiana*, and a dominant growth of shrubs, typical of open habitat in the degraded forest site. Degradation of forests after logging or human-induced activities, especially cutting of trees, grazing practices, and incendiary forest fires often promotes colonization and dominance by perennial shrubs like *Viburnum* species and *Sambucas* species. Increased colonization of these aggressive weedy species may be responsible for affecting germination of seed falls and survival and establishment of primary forest species, due to intense competition for resources in the degraded forest site. Similar concept was given by [22]. Loss of organic matter by water erosion, accelerated oxidation and trampling of grazing animals probably account for the higher compaction, decreased porosity and poor structural stability of soils under the degraded forest site [23], [24]. Soil compaction generally affects growth and development of plant roots and restricts water and nutrient availability to plants [25]. [26].

community is under the high pressure of locals for their greedy requirements. The most vulnerable tree species in Yousmarg forest area is the *Pinus wallichiana* (common name blue-pine) (local name "Kayur").

Table 1
 Species Composition of Forest Vegetation

Herbs			
<i>Acheilia millefolium</i>	<i>Leucanthemum vulgare</i>	<i>Medicago sp.</i>	
<i>Alchemilla sp.</i>	<i>Plantago lanceolata</i>	<i>Rumex sp.</i>	
<i>Barbarea sp.</i>	<i>Poa sp.</i>	<i>Veronica sp.</i>	
<i>Cerastium ceratoides</i>	<i>Trifolium sp.</i>	<i>Lenontopodium sp</i>	
<i>Polygonium sp.</i>	<i>Potentilla sp.</i>	<i>Prunella sp.</i>	<i>Taraxacum sp</i>
<i>Cyanodon dactylon</i>	<i>Polygonum hydropiper</i>	<i>Geranium sp.</i>	<i>Geum sp.</i>
<i>Euthrasia officinale</i>	<i>Thymus sp.</i>	<i>Ranunculus sp.</i>	<i>Fragaria sp.</i>
Shrubs			
<i>Sambuca sp.</i>		<i>Viburnum sp.</i>	
Trees			
<i>Pinus wallichiana</i>		<i>Picea smithana</i>	<i>Abies pindrow</i>

Table 2

Spatial variation in various plant species indices at the two sites protected and grazing area of
 Yousmarg main bowl Nabrepathar

HERBS	Species	Dominance	Species	Evenness	Similarity
AT :	Diversity	(Simpson	Richness(Margal	Index(Pielo	Index
	Index H'	Index)	ef Index)	u Index)	(Sorenson
					Index)
PROTECTED	2.97	0.97	5.75	0.93	57.42
SITE.					
GRAZING	2.33	0.87	3.70	0.88	
SITE.					

Table 3

Comparison of tree species on the basis of Frequency (F), Density (D) and Abundance (A) from protected and degraded forest area

NABREPATHER SITE						
	PROTECRED AREA			DEGRADED AREA		
Species	F	D	A	F	D	A
<i>Pinus wallichiana</i>	75	1.75	2.00	50	0.75	1.50
<i>Picea smithana</i>	100	3.25	3.25	75	1.75	2.33
<i>Abies pindrow</i>	100	2.75	2.75	75	1.5	2.00
MASHID NAR SITE						
	PROTECRED AREA			DEGRADED AREA		
Species	F	D	A	F	D	A
<i>Pinus wallichiana</i>	75	1.25	1.66	50	0.75	1.00
<i>Picea smithana</i>	100	2.75	2.75	75	1.00	1.33
<i>Abies pindrow</i>	100	2.50	2.50	75	1.25	1.66
CHOTA YOUS SITE						
	PROTECRED AREA			DEGRADED AREA		
Species	F	D	A	F	D	A
<i>Pinus wallichiana</i>	100	6.5	6.5	75	1.75	2.33
<i>Picea smithana</i>	100	3.25	3.25	100	2.75	2.75
<i>Abies pindrow</i>	75	1.25	1.66	25	0.25	1.00

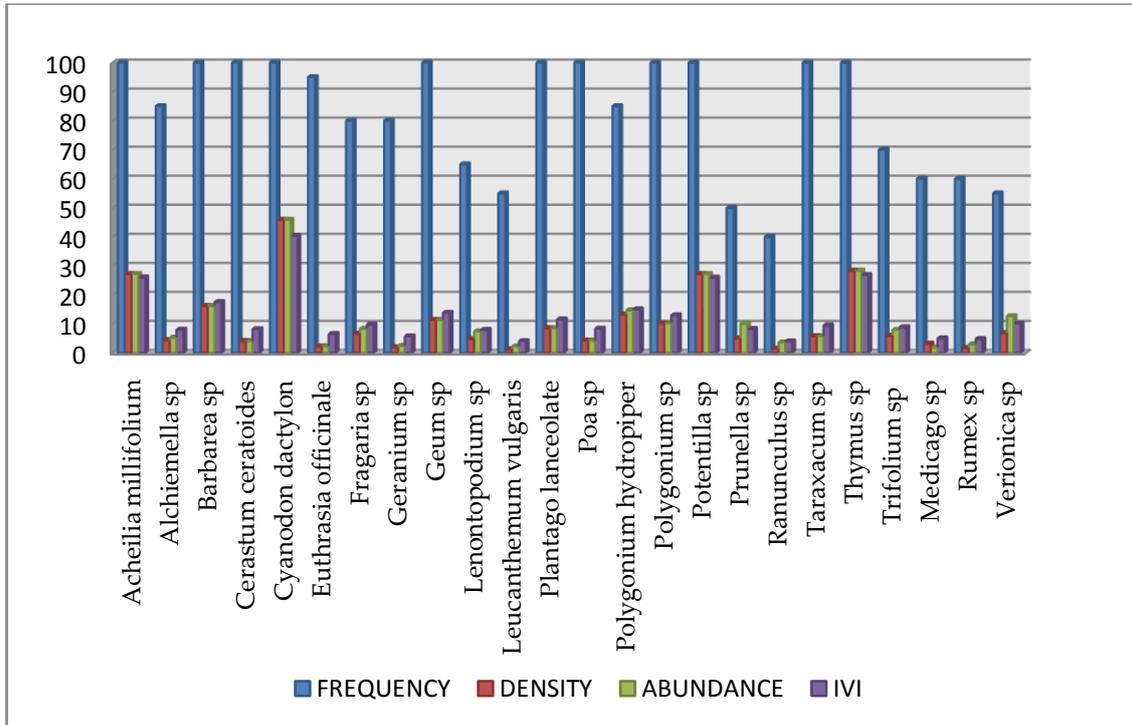


Fig 1: Frequency, density, abundance and IVI of the herbs at fenced site Yousmarg main bowl.

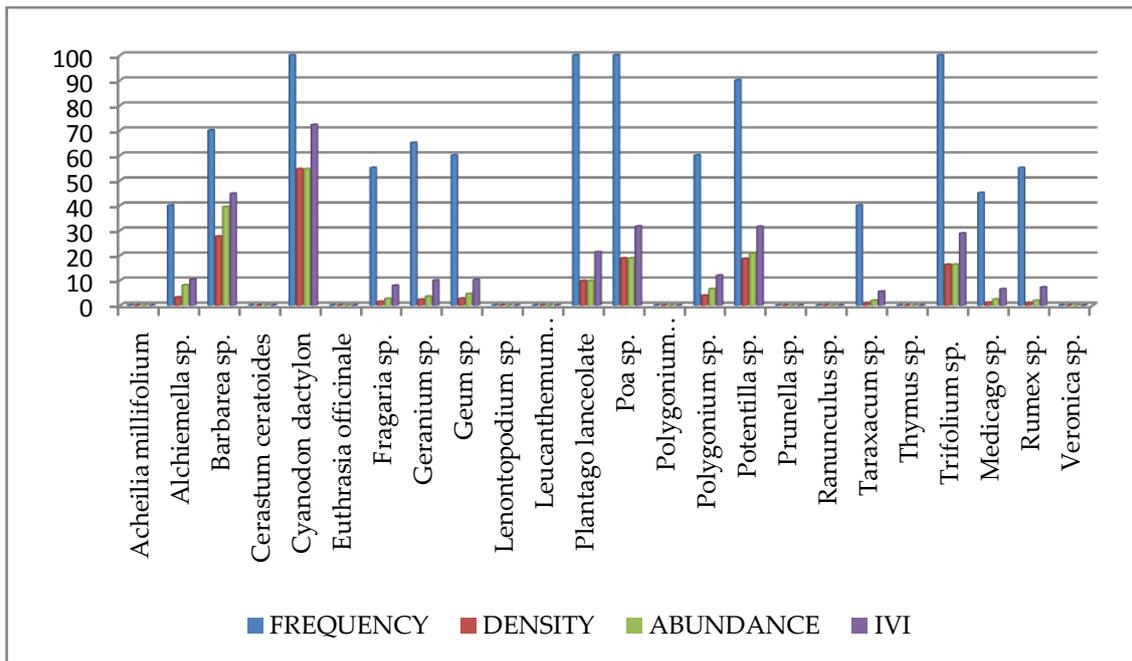


Fig 2: Frequency, density, abundance and IVI of the herbs at grazing site Yousmarg main bowl.

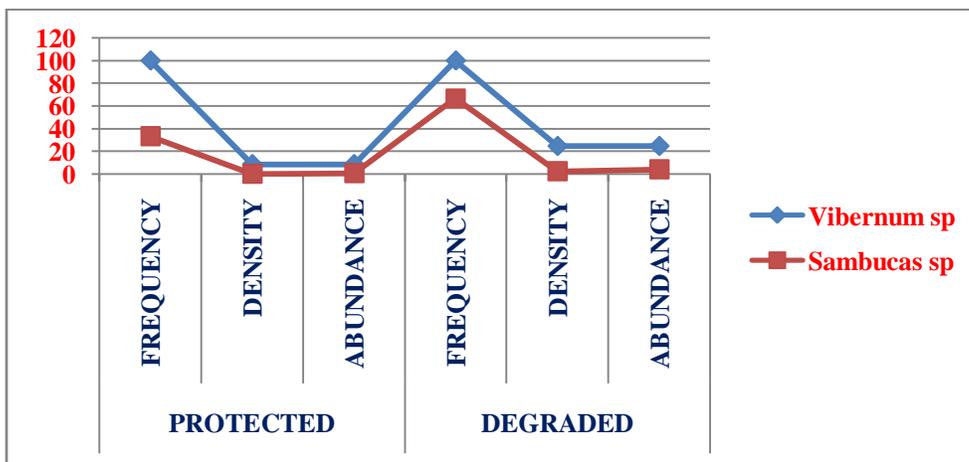


Fig 3: comparison of shrubs at protected site and degraded site on the basis of frequency, density and abundance.

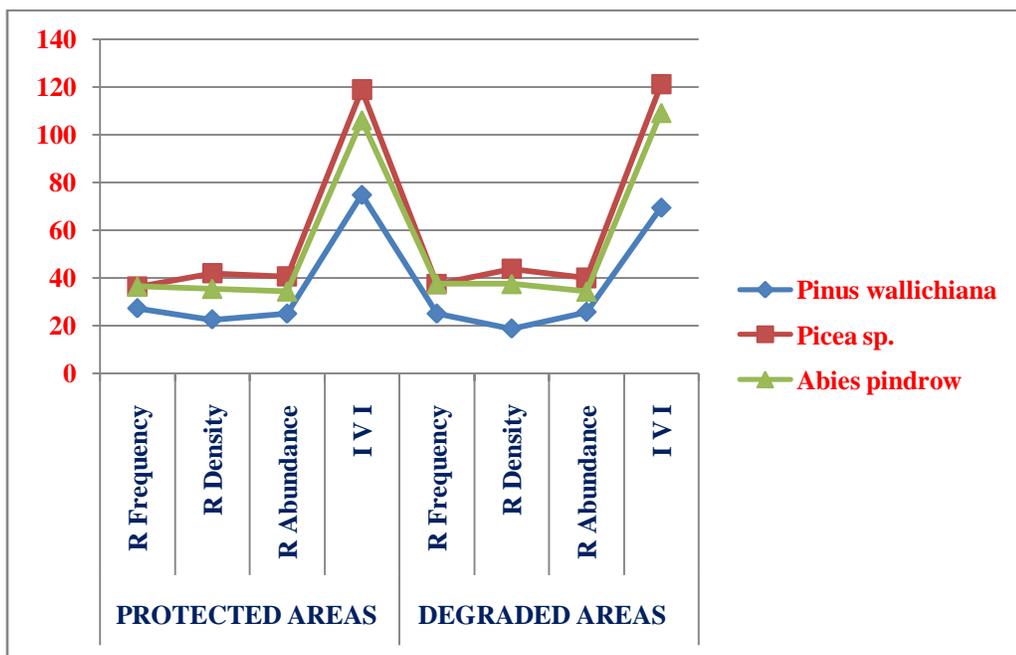


Fig 4: Comparison of tree species on the basis of importance value index (IVI) in protected and degraded forest area, Nabrepathar site

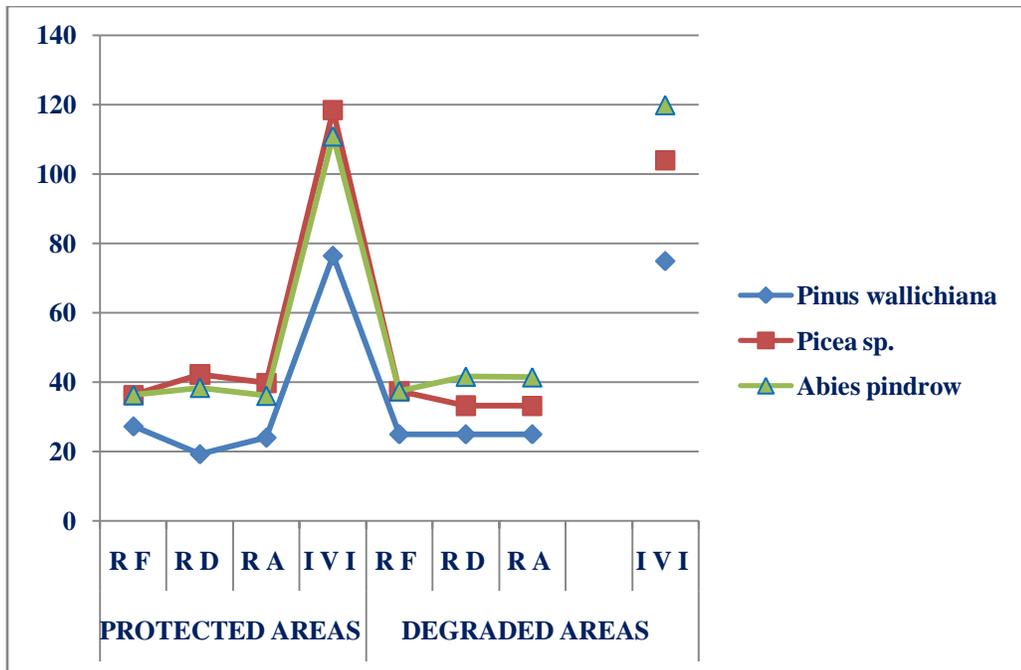


Fig 5: Comparison of tree species on the basis of importance value index (IVI) in protected and degraded forest area, Mashid Nar.

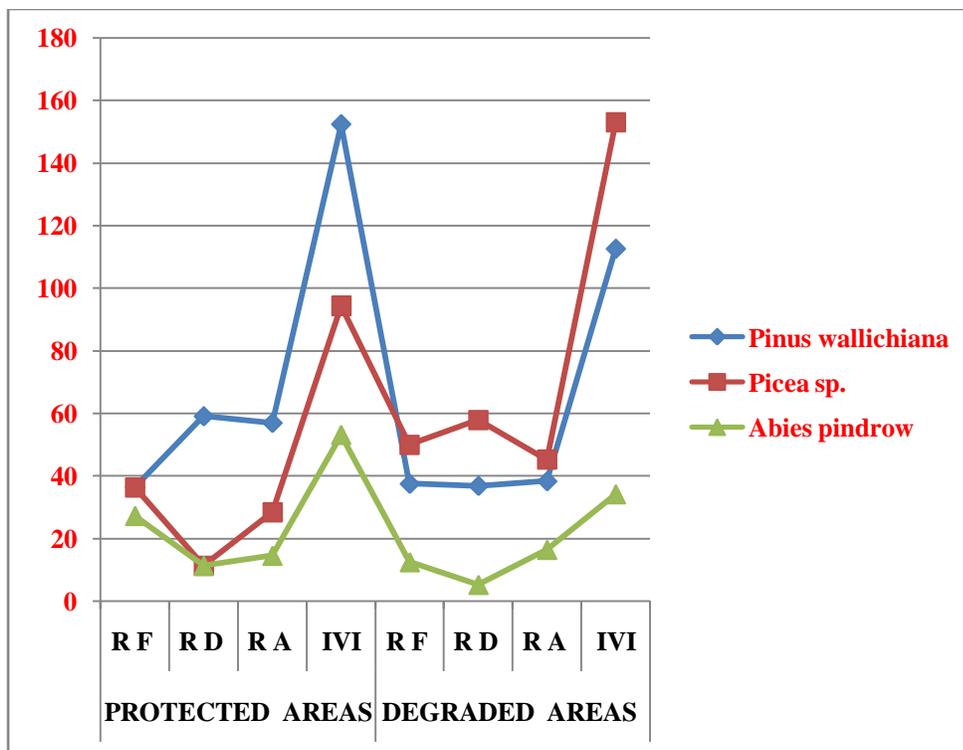


Fig 6: Comparison of tree species on the basis of importance value index (IVI) in protected and degraded forest area at Chota Yous

ACKNOWLEDGMENT

The present work is a part of M.Sc Environmental Science project work-2010-2011, of the first author. Thanks are due to Head, P. G. Department of Environmental Science University of Kashmir for providing necessary laboratory facilities as well as the transportation and also to staff members especially Dr. Samiullah Bhat, Dr. Arshid Jehangir and Dr. Mudasar Ali for their valuable suggestions. Last but not Least, warm thanks to Mr. Idrees Yousuf Dar for helping in formatting this paper.

References

- [1] Mathur, H. N., Singh, R. P and Gupta, M. K. (1992). Comparative Study of infiltration in soil under forest cover and agriculture in temperate climate. *Indian For.*, 108 (10): 648-652
- [2] Lone, H. A. and Pandit, A. K. (2005). Impact of deforestation on some economical important tree species of Langate forest division in Kashmir Himalaya. *J. Res. Dev.*, 5: 39-44.
- [3] West Virginia University Extension Service Overgrazing Can Hurt Environment, Your Pocketbook Ed Rayburn. 2000.
- [4] Misra, R. 1968. *Ecology Workbook*. Oxford and I.B.H. publishing Co., New Delhi.
- [5] Kershaw, K.R. (1973): *Quantitative and Dynamic Plant Ecology*. Edward Arnold Ltd., London.
- [6] Shannon, C. E. and Wiener, W., (1963): *The Mathematical Theory of Communication*. University Press, Illinois, USA.
- [7] Simpson, E.H. (1949): Measurement of diversity. *Nature*, 163: 688.
- [8] Pielou, E.C., (1975): *Ecological Diversity*. John Wiley and Sons. New York
- [9] Margalef, D., R. (1958). Information theory in ecology. *Year book of the Society for General Systems Research*. 3: 36-71.
- [10] Sorenson, T. (1948). A method of establishing groups of equal amplitude in a plant based on similarity of species content. *K. Dan. Vidensk. Selsk.* 5: 1-34
- [11] Vesk, P.A., and Westoby, M. (2000). Predicting plant species responses to grazing. *J. Applied Ecol.*, 38: 897-909.
- [12] Sher, H., Khan, Z.D., Khan, A.U. and Hussain, F. (2005). In situ conservation of some selected medicinal plants of Upper Swat, Pakistan. *J. Acta Botanica Yunnanica*, 27: 27-36.
- [13] Kukshal, S., Nautiyal, B.P., Anthwa I.A., Sharma, A. and Bhatt, A.B., 2006. Phytosociological investigation and life form pattern of grazing lands under pine canopy in temperate zone, Northwest Himalaya, India. *Res. J. Botany*, 4: 55-69.
- [14] Skornik, M. vidrih and M. kaligalic, (2010). The effect of grazing pressure on species richness, composition and productivity in North Adriatic Karst pastures. *Plant biosystems*. 144 (2): 355-364.
- [15] Mushtaq, B. and Pandit, A. K. (2010). Impact of biotic factor on the vegetation of Shankeracharya forest ecosystem. *J. Himalayan Ecol. Sustain. Dev.* 5 (5): 39-44.
- [16] Kumar, M., Sharma C.M. and Rajwar, G.S. 2004. A study on the community structure and diversity of a sub-tropical forest of Garhwal Himalayas. *Indian Forester*, 130 (2): 207-214.
- [17] Rajwar, G.S. and Gupta, S.K., (1992). Structure of forest vegetation of Garhwal Siwalik Hills between Khoh and Ganga. *Indian Forester*, 118 (2): 148-165.
- [18] Mishra, A., Mishra, L., Sharma, C.M. and Sharma, S.D. (2002). Community structure and growing stock variation on different aspects of dry semi-evergreen Shorea *Robusta* forest cover type of Garhwal Himalaya. *Annals of Forestry*, 10(2): 233-242.
- [19] Curtis, J. T. and McIntosh, R. P. 1951. The inter-relation of certain analytic and synthetic Phytosociological character. *Ecol.* 31: 343-455.
- [20] Mishra, R.K., Upadhyay, V. P. and Mohanty. R.C. (2003). Vegetation diversity of Similipal biosphere reserve. *E-Planet.1* (1): 4-9.
- [21] Wali, M. K. (1964). A Preliminary survey of the conifer community of Kashmir Himalaya. *Tropical Ecology*, 5: 32-41.
- [22] Islam, K. R. and Weil R. R., 2001. Land-use effects on soil quality in a tropical forest ecosystem of Bangladesh. *Agriculture, Ecosystems and Environment*, 79: 9-16.
- [23] Aweto A. O., Adejumbi, D. O. 1991. Impact of grazing on soil in the Southern Guinea Savannah Zone of Nigeria. *The Environmentalist*. 11: 27-32.
- [24] Dagar J. C., Mongia A. D., Singh N. T., 1995. Degradation of tropical rain forest soils upon replacement with plantations and arable crops in Andaman and Nicobar Islands in India. *Tropical Ecology*, 36: 89-101.
- [25] Weert, V. A., (1974). The influence of mechanical forest clearing on soil conditions and resulting effects on root growth. *Tropical Agriculture (Trinidad)*, 51: 325-33.
- [26] Chauvel, A., Grimaldi, M. and Tessier, D. 1991. Changes in soil pore-space distribution following deforestation and regeneration: An example from the central Amazon Basin, Brazil. *Forest Ecology and Management*, 38: 259-271.

Matter and Empty Universe

D.I.Sanitsar

We do not know what is the minimal and maximal both side unlimited of our universe system. I suppose in the unlimited of Universe the matter transfers to emptiness. On the contrary from emptiness regenerate new matter.

Overall Structure of the Universe

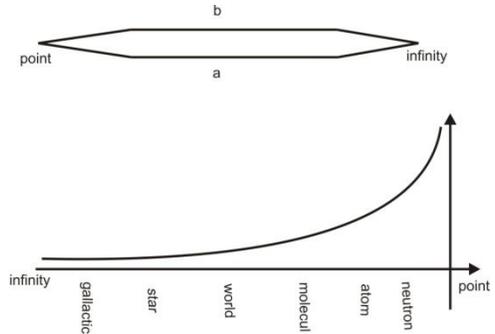
The Universe is created in common algorithm. The law and phenomenon natural have very simple explanation. The micro and macro universes have one genuine origin. In other words the world has one (absolutely) phenomenon of emergence display and affection beyond forties a time impression.

We don't know what electron and core atom is. We don't know neither what kind elementary particle composed the electron, proton and neutron. Only we suppose that it is very small particle. What structure of this elementary particles.

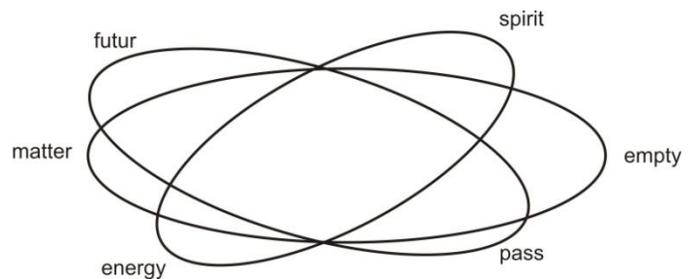
Just there elementary particles emitted energy when syntheses or disintegration atoms. It is energy of super smaller elementary particle. Also we don't know about the minimal and maximal unlimited of our universe.

The very small particles compose elementary particle. The elementary particles compose atom, the atoms compose molecule, the molecules compose sun system, the stars compose galactic, and the galactic compose the universe. So on the contrary we are consider that point is infinitely then the stars compose the atom.

So far we have accepted unanimously that the world is consisted of electrons and protons, but we still don't know what is behind of electrons and atoms (beyond L1), neither we nor Universe our vision can imagine, galaxies etc. So we understand the matter exist is space, energy and time.



So existing the matter exist space, if exist space then curing the time. We are thinking the time is passing unreturned. This is mistake, time returned. But by other way. Such as the water flow streaming down and water steaming evaporate to transferring in cloud, the rain bring get back and the water curing again. As for me the time has its nature of circulation. So I understand there must be in separable inter connection between the matter is it is existing in time. The time is we understand it must have its past. So there must be a future. Only in the material Universe the time curing from passes to future. In the emptiness the future transferred to pass.

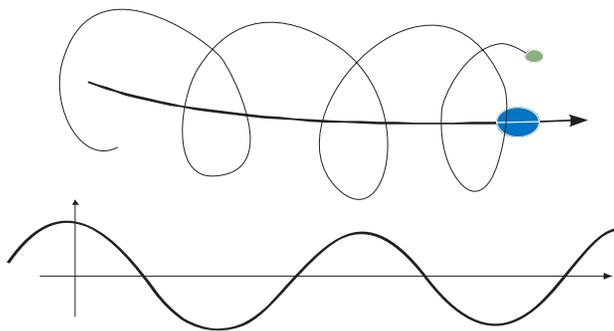


Sense the mechanism **Big bang** is reflection of the moment future to past transformation. As we understand, these days, due to Big Bang the Universe is spreading to all direction and dimensions (beyond L2). The terming Big

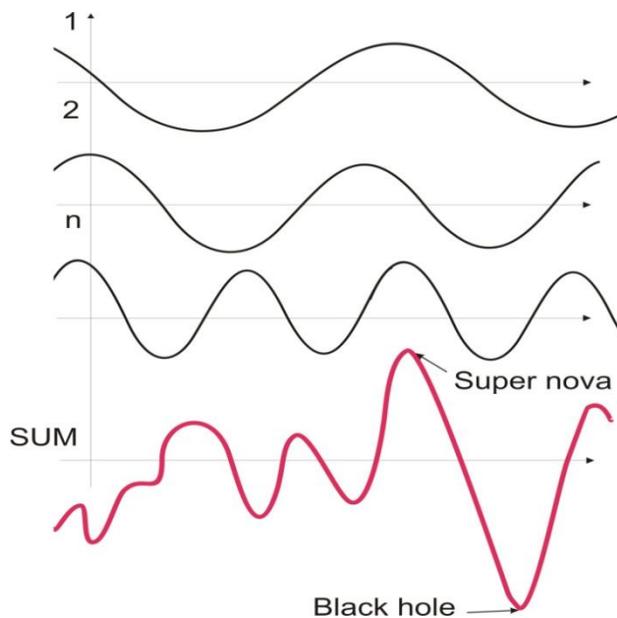
• D.I.Sanitsar is mathematician graduated from Moscow GP University, researcher clean technology developer. Email: sanitsar@gmail.com

Bang is understand today as Natural cosmic disaster and we humanity is still on the way directed by this mechanism. In my view the mechanism of the Universe in our vision was not provoct by one and only such an extreme blow. The sense of existence of our Universe innumerable and continues such explosions. Therefore you can understanding that **PASS=FUTURE**.

All the universe is consist over the rotary motion. Therefore some event from time to time return at regular reiterated. Such bodies the electrons or the world are said to have a motion of vibration or oscillation.



Every rotary mechanism has its physical impact: waves. Other words the waves born from matter circulation. From general physic the waves have inter affection positive and negative.

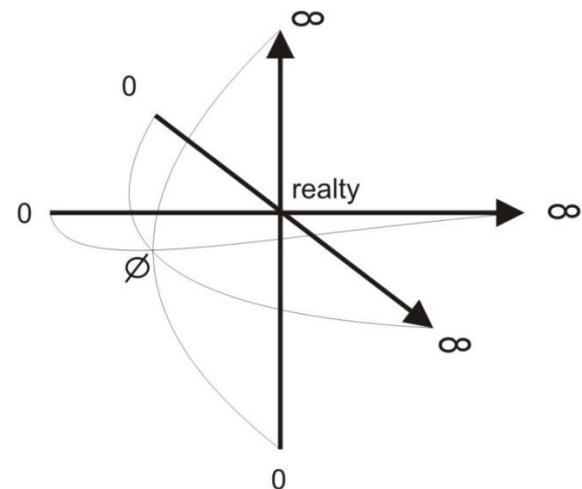


If we shall add different waves over each other the result shall produce diffraction. This kind of interference leading shape the diffraction. Rule of diffraction mechanism is

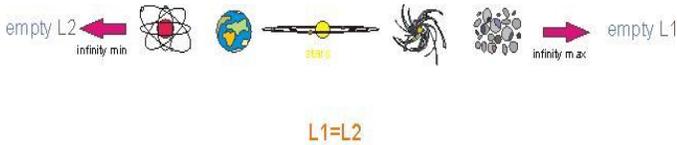
under law subtraction and summation waves. We are called this diffraction named **black hole** and antipode **supernova**.

The universe is existing too many years eternally, everlasting. You see only short time interval of the existence. Every material substance has its mass, speed, inertia, and circulation and rotary motion mechanism. Black hole is not emptiness; this is only one of result diffraction of interference stars motion energy. Therefore black hole changing here disposition in long time.

And my question as when and where its shall end. I suppose the Universe unlimited; therefore the energy atoms stars, galaxies have shall reduce and disappear at the end of their tripe. In my vision this energy shall transferred to spiritual energy. In other words the sense of spiritual energy the existence it's into empty realty. Thus, we have arrived to the point when where no time, no mass is and neither no space, in other words there shall be they absolute Neil.



The smallest particle forming an atom or the smallest particle, which we cannot define, can be reduced without limit till it is dissolved into emptiness. A galaxy, which is a concentration of stars, as well as the re-grouping or concentration of galaxies also turn into emptiness with the reduction of energy.



The emptiness of unlimited macro-universe L1 and the emptiness of unlimited micro-universe L2 are dissolved into each other, and constitute the only eternally connected thing. In other words, L1=L2 (L1 is equivalent to L2).

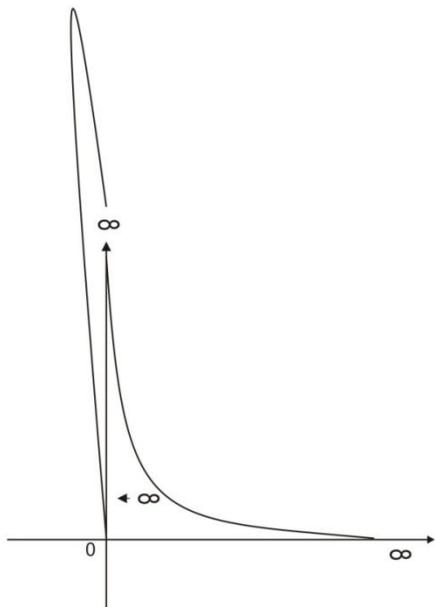
Therefore

$$\frac{1}{0} = \infty \quad \frac{1}{\infty} = 0$$

$$\text{If } \lim \infty = \lim 0 = \emptyset$$

$$\text{Then } 1 = \infty * 0 = \emptyset * \emptyset = \emptyset^2$$

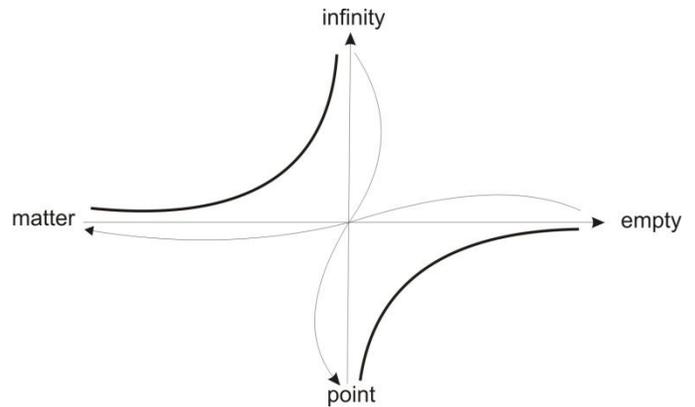
If multiply emptiness by empty we get one unit of i.e. matter.



At the unlimited of universe, the energy is exhausted and transferred to spirit. By the principle of Einstein which shows the equivalence of mass and energy, the mass will disappear and motion will stop. The matter transfers to emptiness. This is Emptiness of the emptiness. Please don't confuse between emptiness and vacuum. So I mine vacuum is absolutely different of Neil. The vacuums have space

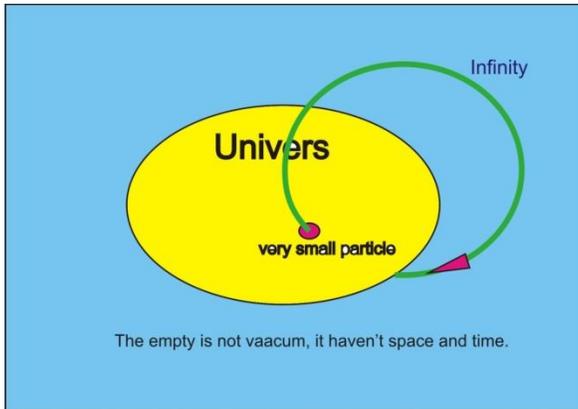
and time, in emptiness hasn't space, energy, time and matter. If we stop all the motions of universe, the universe transferred also to empty. The space transform a nothing, the time transform nonsense. Some time our earth rotation motion has stopped but not definitely. Example of this occurrence is event of recent braking core of our earth. This time all mantle pressed from west to east by here inertia and create mountains flexure in east side each continents. May be this was Noah's Ark time, rise great tsunami and cover all dry land destroy life. Clarity trace of this great tsunami leaves in Gobi Mongolia.

Actually the phenomenon of absolute Neil or Emptiness as both beyond (Beyond L1=Beyond L2) is the very sense of the existence of the Universe. Our spirit connected with empty universe by specially mechanism of God. When infinitely transferred point, in empty spring up rotary moment under the spirit action this moment get energy.

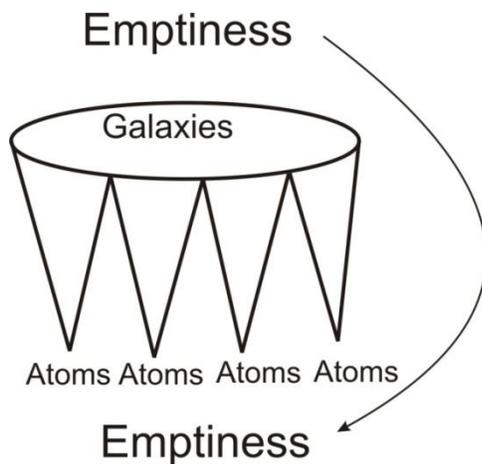


The very sense of Natural is rotation any kind of energy in material suspense. Due to this mechanism or law of order as soon as material energy shall develop (or transfer) from spiritual energy they strength (scope and scale) it shall gain power and motion, leading to the standard existence of material.

technique we were able to design is a break capability a human being could ever produce.



Energy given mass and from empty appears matter. This matters compose super micro elementary particle, the super micro elementary particles composed neutron. Then constructed the space, and appear positively and negatively charged particles. Then compose atom. Transform unlimited the matter to emptiness and from emptiness regenerates new super smallest matter.



As soon as the matter comes in to existence shall be under Law of Nature rules. There are unimagined versions of matter following destine. This destine shall depend under the law of time. Sense of law is general for whole of Nature. Every kind of matter shall have its lifespan. But where is end of existence (life and time). So sense something exists (birth present, past and future). Let's talk about sense of Future. The sense future is the moment.

Every time reascent billion and billion new very small particles and disappearing billion and billion galaxies. The Universe we are able to catch with our eyes and the

Removal of Baseline wander and detection of QRS complex using wavelets

Nilesh Parihar, Dr. V. S. Chouhan

Abstract — ECG signals are used to detect the heart rate and heart abnormalities. For extraction of ECG features and detection of QRS complexes it is required to remove baseline wander and minimize the noise interference. In this paper we proposed a technique to remove baseline wander using Kaiser Windowing filter and wavelet transform, among which wavelet is most powerful and effective tool for analyzing transient signal. The algorithm is developed in matlab with standard CSEECG database.

Index Terms— Baseline wander, Differentiator, ECG, Kaiser Window, MAI, QRS Complex, Wavelets.

1 Introduction

ECG provides useful information about function status of the heart beat. This technique is great importance for the analysis & detection of the heart cardiac anomalies. In most of the practical situations the ECG signal is available for observation, contains significant drift from the baseline or zero line of the ECG plot. The ECG signal captured by a computerized ECG machine can be rendered to the algorithm suggested in this paper and a clean drift-free signal would be available for visual inspection by a cardiac specialist as well as for computerized detection of the ECG wave complexes. [7]

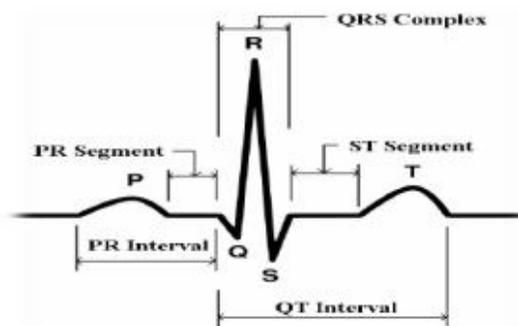


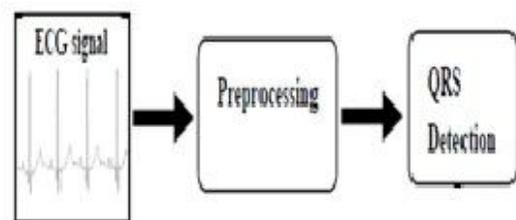
Fig.1. QRS Complex

A standard scalar ECG is shown in fig.1. It consists of P-wave, PR-interval, PR-segment, QRS complex, ST-segment, ST- interval and T-wave. The P wave represents atrial depolarization, the QRS-complex represents left ventricular depolarization and the T-wave represents the left ventricular depolarization.

A common technique utilized in the QRS detector algorithm is to employ a scheme that consists of a pre-processor and a decision rule. The purpose of the pre-processor is to enhance the QRS, while suppressing the other complexes as well as the noise and the artifacts. The pre-processor consists of a filter, differentiator and moving average integrator and the purpose of the decision rule is to determine whether or not QRS complex is present at a given instant in the signal. This paper covers the ECG signals pre processing and implementation of wavelet dB6 on pre-processor signals. Then result is implemented on prepared algorithm. The algorithm is tested on standard CSE ECG database, having multi-lead ECG signal recordings for 125 cases. Each digital record constitutes a 10 second recording containing 5000 samples taken at a rate of 500 samples/sec.

2 Analysis Method

In order to extract information from the ECG signal, the ECG signal is divided into two steps: Pre processing and Feature Extraction as shown in fig. 2.



- Nilesh parihar is currently pursuing PhD program in digital communication from jodhpur national University, India . He is currently working as faculty of Bhartiya Institute of engineering and technology, sikar.PH-+91-9929118025. E-mail:parihar_nilesh@rediffmail.com
- Dr. Vijay S. chouhan is currently an associate professor in M.B.M. engineering college, India.

Fig.2. ECG Signal Processing

The Pre processing stage removes noise from the ECG signal by using filtering method and Feature Extraction is performed by using Discrete Wavelet Transform (DWT) dB6.[2]

In pre-processing, ECG signal mainly contains noises of different types, namely frequency interference, baseline drift, electrode contact noise, polarization noise, muscle noise, the internal amplifier noise and motor artifacts. Artifacts are the noise induced to ECG signals that result from movements of electrodes. For removing the baseline wander and noise we design and follow a steps as a band pass filter with the use of Kaiser Windowing technique with cut off frequency 0.5-40 Hz, differentiation, squaring and moving average integrator [3]. After removing baseline wandering, the resulting ECG signal is more stationary.

Feature extraction is to obtain our wavelet analysis; we used the Matlab program, which contains a "wavelet toolbox". The WT uses a short time interval for evaluating higher frequencies and a long time interval for lower frequencies. Wavelet Transform of a signal $f(t)$ is defined as the sum of over all time of the signal multiplied by scaled, shifted versions of the wavelet function Ψ and is given by,

$$W(a, b) = \int_{-\infty}^{\infty} f(t) \Psi_{a,b}(t) dt$$

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi^* \left(\frac{t-b}{a} \right)$$

Where * denotes complex conjugation and, $\Psi_{a,b}$ is a window function called the mother wavelet, 'a' is a scale factor and 'b' is a translation factor. Here is a $\Psi \left(\frac{t-b}{a} \right)$ shifted and scaled version of a mother wavelet which is used as bases for wavelet decomposition of the input signal. If the scale parameter is the set of Integral powers of 2, i.e., $a = 2^j$ ($j \in \mathbb{Z}$, \mathbb{Z} is Integer set), then the wavelet is called a dyadic wavelet [12]. The Wavelet Transform at scale 2^j is given by

$$Wf(2^j, \tau) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} f(t) \Psi \left(\frac{t-\tau}{2^j} \right) dt$$

We define local maxima of the Wavelet Transform modulus [13] as: - Let $Wf(x)$ is the Wavelet Transform of a function $f(x)$;

1. We call a local extreme any point x_0 such that $d(Wf(x)) / dx$ has a zero crossing at $x = x_0$, when x varies.
2. We call a modulus maximum; any point x_0 such that $|Wf(x)| < |Wf(x_0)|$ when x belongs to either a right or left neighbourhood of x_0 , and $|Wf(x)| \leq |Wf(x_0)|$ when x belongs to the other side of the neighbourhood of x_0 .

3. We call maxima line, any connected curve in the scale space x along which all points are modulus maxima.

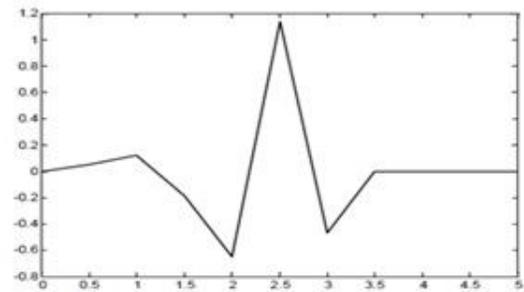


Fig.3. Daubechies db6 wavelet

In Fig. 3 is presented a 3-level signal decomposition of a sample ECG waveform using the db6 wavelet. This first decomposes the ECG signal into several sub bands by applying the Wavelet Transform, and then modifies each wavelet coefficient by applying a threshold function. The high frequency components of the ECG signal decreases as lower details are removed from the original signal. As the lower details are removed, the signal becomes smoother and the noises on the T and P waves disappears since noises are marked by high frequency components picked up along the ways of transmission.

3 QRS complex Detection

This is the contribution of the discrete wavelet transform where noise filtration is performed implicitly. In Fig.4 is presented the ECG signal before and after noise removal. The detection of the QRS complex is the last step of feature extraction. The R peaks have the largest amplitudes among all the waves, making them the easiest way to detect and good reference points for future detections.

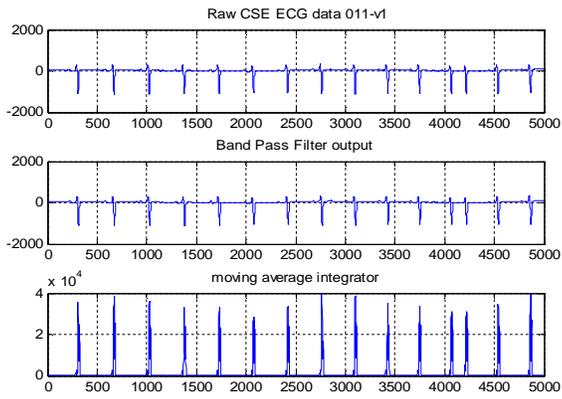
The signal was processed using the wavelets up to 6 levels. However for the detection of the QRS complex, only details up to level 3 were kept and all the rest removed. This procedure removed lower frequencies considering QRS waves have comparatively higher frequency than other waves [6].

A threshold condition is the maximum value which is sub sequentially applied to set a practical lower limit to help to remove the unrelated noisy peaks. At this point, the data set is ready for peak detection through a very simple search algorithm that produces very accurate results. In Fig. 5 is presented the ECG signal and the extracted peaks corresponding to the QRS complex.

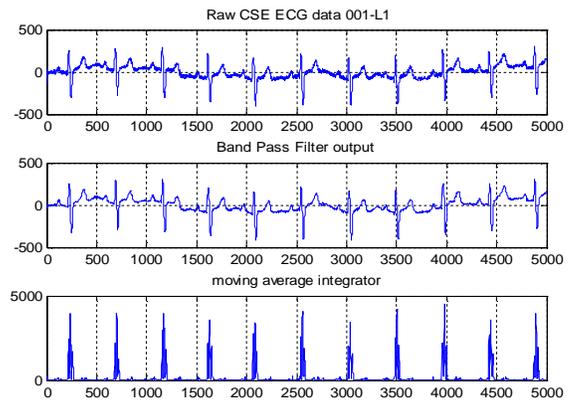
4 Results

The results obtained with the proposed wavelets using different ECG noise levels. The algorithm based on wavelet for the detection of QRS complex of ECG signal. It can be notice that good results obtained with the db6 wavelet are

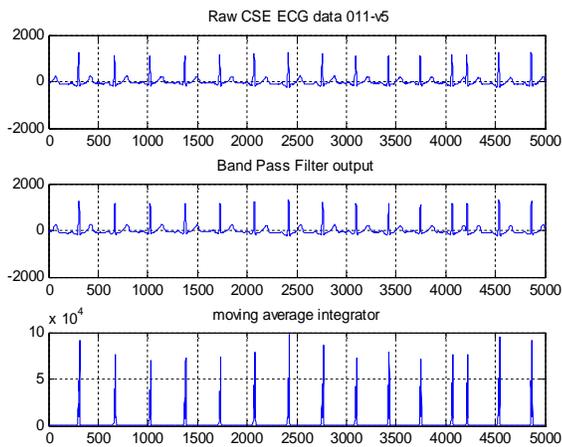
caused by the resemblance that exists between this wavelet and the actual ECG signal.



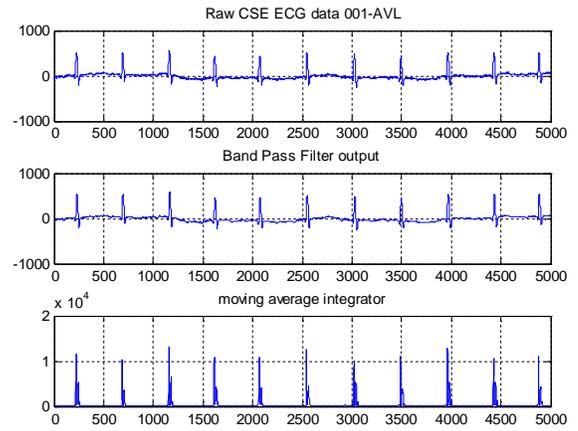
(a)



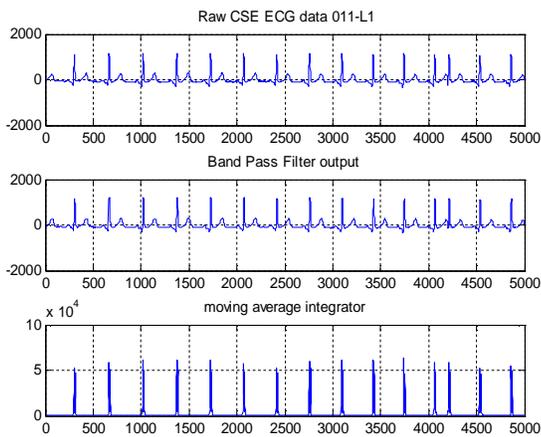
(d)



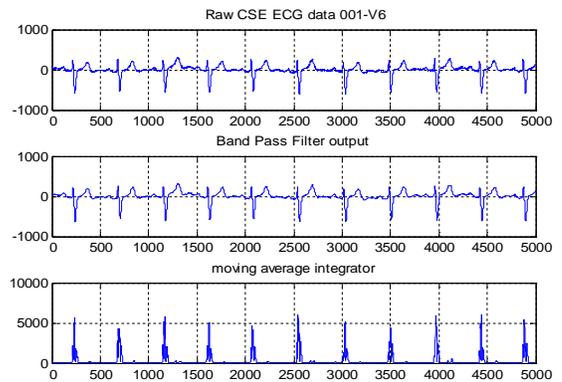
(b)



(e)



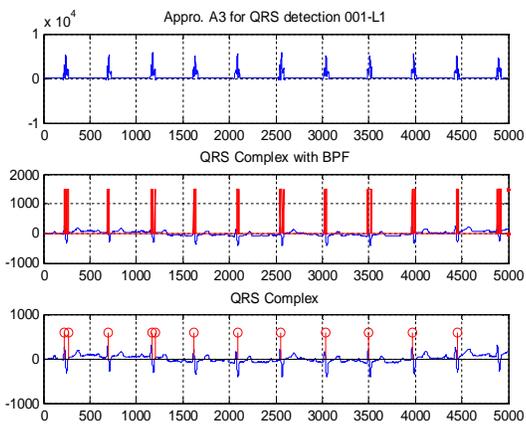
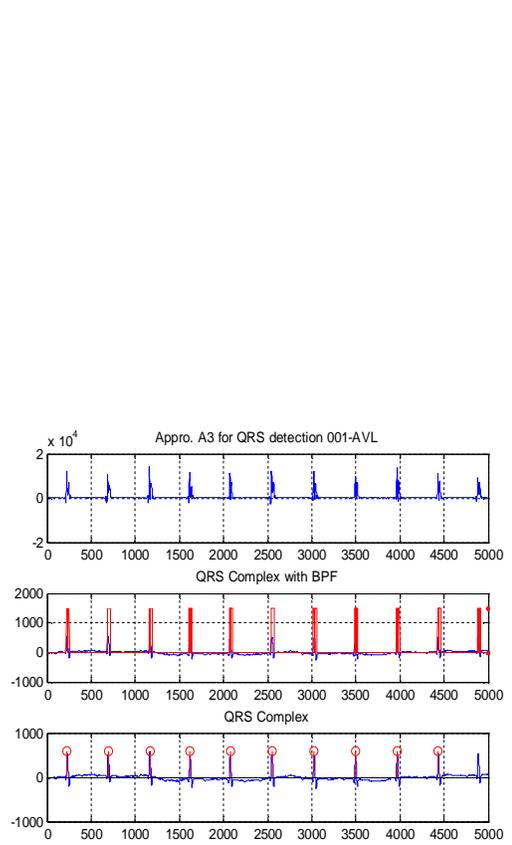
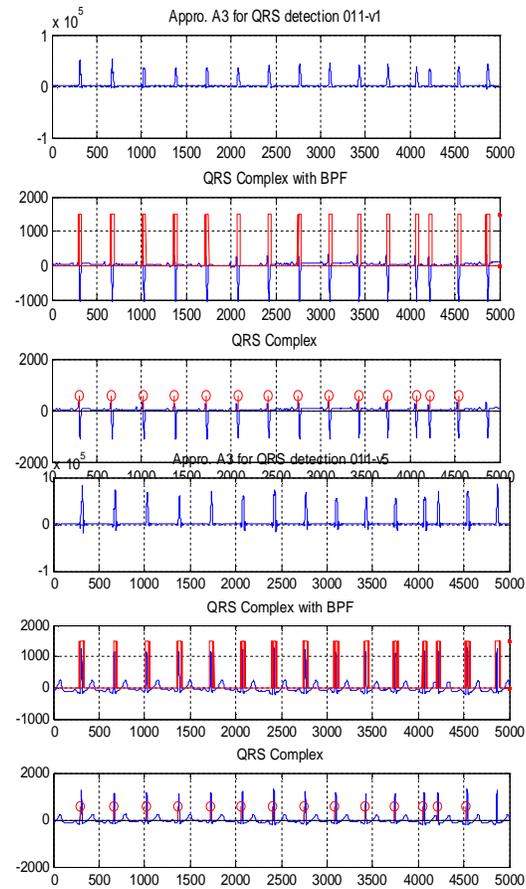
(c)



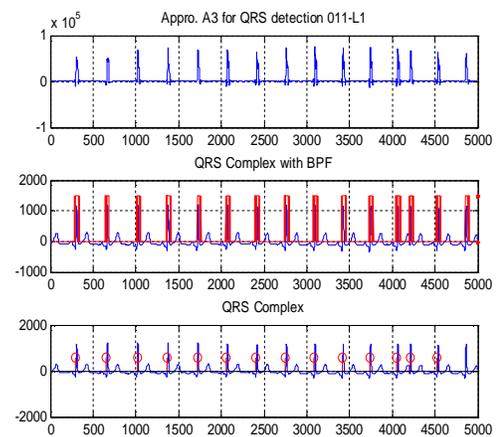
(f)

Fig.4(a)(b)(c)(d)(e)(f) inputs with their corresponding filtered outputs

(a)



(e)



(

d)

(b)

(c)

(f)

Fig.5 (a)(b)(c)(d)(e)(f) R peaks defined using wavelet

REFERENCES

- [1] J. PAN AND W. J. TOMPKINS, "A REAL-TIME QRS DETECTION ALGORITHM", IEEE TRANS. BIOMED. ENG., VOL. 32, PP. 230-236, 1985.
- [2] P. SASHI KALA, DR. R.S.D. WAHIDA BANU, "EXTRACTION OF P WAVE AND T WAVE IN ELECTROCARDIOGRAM USING WAVELET TRANSFORM", (IJCSIT) INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGIES, VOL. 2 (1) , 2011, 489-493.
- [3] S. S. Mehta, and N. S. Lingayat, Member, IAENG, "Support Vector Machine for Cardiac Beat Detection in Single Lead Electrocardiogram", IAENG International Journal of Applied Mathematics, IJAM.
- [4] LG Herrera-Bendezd, J Garcia, BG Denys, "Real-Time Digital Filters for ECG Signals: Evaluation and New Designs, 0-8186-2485-X/92, 1992 EEE.
- [5] Gordan Cornelia, Reiz Romulus, ECG Signals Processing Using Wavelets, ersity of Oradea: Electronics Department, Faculty of Electrical Engineering and Information Technology, Oradea, Romania.
- [6] ZHI-DONG ZHAO, Y. QUAN CHEN, "A New Method for Removal of baseline wonder & Power Line interference in ECG signals", proceeding in 5th International conference on Machine Learning & Cybernetics, Dalian, 13 - 16 August 2006.
- [7] Chui K.: "An Introduction to Wavelets", Academic Press, Inc.1992.
- [8] Saritha, V. Sukanya, and Y. Narasimha Murthy, "ECG Signal Analysis Using Wavelet Transforms", Bulgarian Journal of Physics, vol. 35, pp. 68-77, 2008.
- [9] V.S.Chouhan, "Total removal of baseline drift from ECG signals", proceedings of the international conference on computing theory and applications, 2007.

Trust Vs Complexity of E-Commerce Sites

Devendera Agarwal, R.P.Agarwal, J.B.Singh, S.P.Tripathi

Abstract— E-Commerce suffers from uncertainty which can produce devastating results. The user first checks the level of security and then proceeds further. At the same time the user switches to another e-commerce site if he has to deal with several layers of security. To overcome this drawback e-commerce sites are now finding a solution of maintaining high security (Trust) with lesser complexity as far as possible. Our paper focuses on the issue of development of a framework to provide an optimal relationship between the two.

Index Terms— Complexity, Threat to e-commerce, Fuzzy Rule, Security, Tradeoff, Transaction, Trust.

1 INTRODUCTION

INDIA today is facing with various kinds of threat to e-commerce systems. The problem arises when we increase the security of the e-commerce website, the complexity at the user level also increases, which in turn affects the volume of sale. While traditional marketing does not involve any type of complexity since the consumer deals directly with the supplier. Since internet marketing does not involve any face to face direct interaction so a visual interface is essential. There are various types of online buying behavior models like Bettman (1979) and Booms (1981) in which the focus was on personal characteristics viz. Culture, Social Group and Physiological Behavior. Lewis and Lewis (1997) identified five different types of web which remain valid today:

- **Directed information-seekers:** These users will be looking for product information and are not typically planning to buy online.
- **Undirected information-seekers:** These are the users, usually referred to as 'surfers', who like to browse and change sites by following hyperlinks. Members of this group tend to be novice users and may also click banners of the website.
- **Directed buyers:** These buyers are online to purchase specific products online. For such users, brokers that compare product features and prices will be important locations to visit.
- **Bargain hunters:** These users want to find the offers available from sales promotions such as free samples or competitions.
- **Entertainment seekers:** These are users looking to interact with the Web for enjoyment through entering contests such as quizzes, puzzles or interactive multi-player games.

Under all the above categories the main focus is the trust of web users [24], [25] which will finally lead to purchase. The communication between server and client are not secure un-

less it is providing a safe and secure transaction. To reduce the risk we have to deal with development of trustworthiness of the web services, which finally means increasing the complexity of the website.

Trust Level	HIGH	If TRUST level is HIGH and COMPLEXITY is LOW consumer will PREFER this.	If TRUST level is HIGH and COMPLEXITY is HIGH consumer will RESIGN .
	LOW	If TRUST level is LOW then COMPLEXITY has to be LOW consumer will AVOID this.	If TRUST level is LOW and COMPLEXITY is HIGH consumer will ABANDON .
		LOW	HIGH
		Complexity of Transaction	

Fig. 1. Trust/ Complexity Matrix.

From the above figure we can conclude that there has to be some situation in which a trade off between Trust Level and Complexity of the transaction has to be maintained. This trade off can be achieved by the help of development of Fuzzy Rule base, but simple Fuzzy Rule base will not be sufficient for this purpose, so we extend this problem and solve it using Evolutionary Multi-objective Optimism [9], [10], [12].

2 LITERATURE SURVEY

The work by H. Ishibuchi & H. Tanaka (1994) highlights the construction of Fuzzy Classification of various entities using genetic Algorithms. Later on they extended their work (1995) using If-Then-Else rules. M. Setnes (1998) developed a Rule-Based system for developing the Precision & Transparency. D. Nauck (1999) worked on the interpretability aspect of Medical Data; we are motivated by their work and extending it for e-commerce websites. Y. Jin (2000) developed a framework for modeling high dimensional system in finding out their Complexity and Interpretability aspect. L.Castillo (2001) developed the best rule in a genetic fuzzy learning algorithm. M.Setnes (2000) also developed a mechanism of GA-based Modeling & Classification which measures the Complexity and Perfor-

• Devendera Agarwal is currently pursuing his Ph.D. from Shobhit University, Meerut, India. E-mail: dev_goel@in.com
• Prof. R.P.Agarwal & Prof. J.B.Singh, Shobhit University, Meerut, India, E-mail: prajanaq@gmail.com, jbs.tomar@shobhituniversity.ac.in
• Prof. S.P. Tripathi, Department of Computer Science, I.E.T, Lucknow, India, E-mail: spt@ietlucknow.edu

mance of the system.

3 OUR PROPOSED MODEL

Genetic Algorithms [6], [8], [17] have been frequently used to model a solution for conflicting goals. Let Trust (T) be a measure of security which the customer will be provided and Inverse of Complexity (C) be the user comfort level. Applying the Fuzzy Rule base we can get

$$\text{Maximize Trust (T)} \quad (1)$$

But it leads to compromise in the complexity (C) of fuzzy rule based systems [2], [3], [5], [7]. According to consumers survey most of consumers in India considers Trust and Ease of Use (Lower level of Complexity) at the same time. The above problem can be formulated as

$$\begin{aligned} &\text{Maximize Trust (T) subject to} \\ &\text{Inverse complexity (C)} \quad (2) \end{aligned}$$

where complexity (C) is the measure of fuzzy rule system.

We can develop a single objective function to the above solution given as:

$$\begin{aligned} &\text{Maximize } f(\text{Trust (T)}, \\ &\text{Inverse of Complexity (C)}) \quad (3) \end{aligned}$$

We can also use weights in order to determine the exact function for e-commerce site.

$$\begin{aligned} &\text{Maximize } (w_1) \text{ Trust (T)} + \\ &(w_2) \text{ Inverse of Complexity (C)} \quad (4) \end{aligned}$$

We proceed with development of more refined stages in which we can focus on various stages of membership functions. Consider a simple single output function $y = f(x)$ an application of Takagi-Sugeno method [7], [11], [15] we can write it as:

$$\begin{aligned} &\text{Rule } R_1 : \text{ if } x \text{ is } A_1 \text{ then } y = a_1 + b_1x, i = 1, 2, \dots, N \\ &\text{Rule } R_k : \text{ if } x \text{ is } A_k \text{ then } y = a_k + b_kx, k = 1, 2, \dots, N \\ &\vdots \\ &\text{Rule } R_z : \text{ if } x \text{ is } A_z \text{ then } y = a_z + b_zx, z = 1, 2, \dots, N \quad (5) \end{aligned}$$

This output value is given as:

$$y(x) = \frac{\sum_{i=1}^N (a_i + b_x) \mu_{A_i}(x)}{\sum_{i=1}^N \mu_{A_i}(x)} \quad (6)$$

where $y(x)$ is the estimated output value for the input value x and $\mu_{A_i}(x)$ is the membership value of the antecedent fuzzy set A_i .

From the input-output data we can derive the relationship between Trust and Complexity of the e-commerce site considering three Sugeno Rules.

We develop a heuristic rule [1], [2] denoted by three lines A, B and C as the subsequent of the linear function with fuzzy sets A1, A2 and A3. Each of the Fuzzy Rule can be represented in triangular Fuzzy Sets.

Rule R1: If TRUST is SMALL and COMPLEXITY is HIGH Then User's Ease of Use is MEDIUM.

Rule R2: If TRUST is LARGE and COMPLEXITY is MEDIUM Then Users Ease of Use is HIGH

Rule R3: If TRUST is SMALL and COMPLEXITY is SMALL Then Users Ease of Use is HIGH

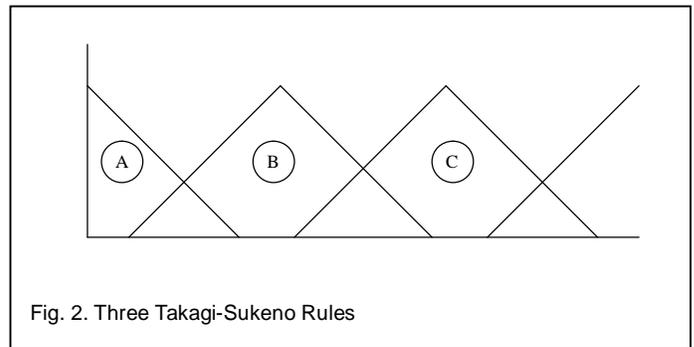


Fig. 2. Three Takagi-Sugeno Rules

Based on the above rules we try to develop a plot between Complexity and Trust and develop our interpretable solution [13], [14], [16], [18], [19] between the two entities.

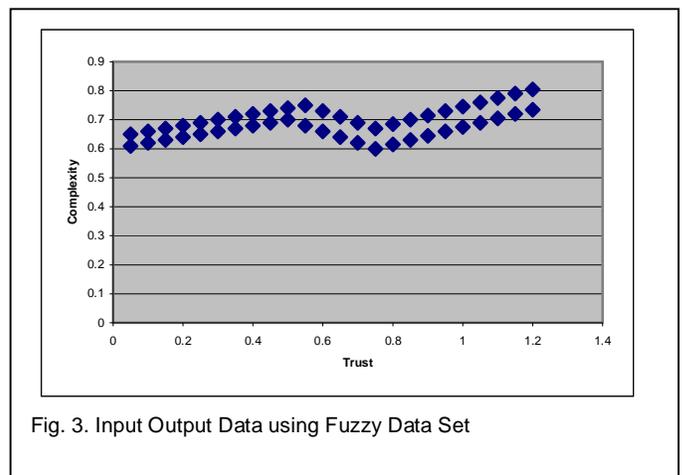


Fig. 3. Input Output Data using Fuzzy Data Set

Possibly we can also merge the above set of rules to achieve more refined results, but a relationship generated by optimization rules gives some gridlines in the area of relationship between the two entities.

4 CONCLUSION

It is very difficult to interpret the exact relationship between the two entities. Different Fuzzy rule are being applied in order to determine the appropriate interpretability. The method that we have used is the application of Fuzzy Optimization Theory [20], [21], [22], [23] to find the probable relationship between Complexity and Trust. The future extension would be to use Evolutionary Algorithm [4], [5] in finding out the best possible trade-off between the two entities.

REFERENCES

- [1] I. Rojas, H. Pomares, J. Ortega, A. Prieto, "Self-organized Fuzzy System Generation from Training Examples", *IEEE Transactions on Fuzzy Systems* 8 (1) (2000) 23–36.
- [2] T. Takagi, M. Sugeno, "Fuzzy Identification of Systems and its Applications to Modeling and Control", *IEEE Transaction Systems, Man and Cybernetics* 15 (1) (1985) 116–132.
- [3] J.C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York, 1981.
- [4] L.A. Zadeh, "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes", *IEEE Transactions on Systems, Man, and Cybernetics* 3 (1973) 28–44.
- [5] J.-S.R. Jang, C. -T. Sun, E. Mizutani, "Neuro-Fuzzy and Soft Computing", Prentice-Hall, Englewood Cliffs, NJ, 1997.
- [6] O. Cordon, F. Herrera, F. Hoffmann, L. Magdalena, "Genetic Fuzzy Systems – Evolutionary Tuning and Learning of Fuzzy Knowledge Bases", World Scientific, Singapore, 2001.
- [7] Y. Jin, W. von Seelen, B. Sendhoff, "On Generating FC3 Fuzzy Rule Systems from Data Using Evolution Strategies", *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics* 29 (6) (1999) 829–845.
- [8] H. Ishibuchi, Y. Nojima, "Analysis of Interpretability–Accuracy Tradeoff of Fuzzy Systems by Multiobjective Fuzzy Genetics-Based Machine Learning", *International Journal of Approximate Reasoning* 44 (1) (2007) 4–31.
- [9] M.J. Gacto, R. Alcalá, F. Herrera, "Adaptation and Application of Multiobjective Evolutionary Algorithms for Rule Reduction and Parameter Tuning of Fuzzy Rule Based Systems", *Soft Computing* 13 (5) (2009) 419–436.
- [10] R. Alcalá, P. Ducange, F. Herrera, B. Lazzerini, F. Marcelloni, "A Multiobjective Evolutionary Approach to Concurrently Learn Rule and Databases of Linguistic Fuzzy-Rule-Based Systems", *IEEE Transactions on Fuzzy Systems* 17 (5) (2009) 1106–1122.
- [11] T.A. Johansen, R. Babuska, "Multiobjective Identification of Takagi-Sugeno Fuzzy Models", *IEEE Transactions on Fuzzy Systems* 11 (6) (2003) 847–860.
- [12] H. Wang, S. Kwong, Y. Jin, W. Wei, K.F. Man, "Agent-Based Evolutionary Approach for Interpretable Rule-Based Knowledge Extraction", *IEEE Transactions on Systems Man, and Cybernetics – Part C: Applications and Reviews* 35 (2) (2005) 143–155.
- [13] Y. Yoshinari, W. Pedrycz, K. Hirota, "Construction of Fuzzy Models Through Clustering Techniques", *Fuzzy Sets and Systems* 54 (1993) 157–165.
- [14] H. Ishibuchi, T. Yamamoto, T. Nakashima, "Fuzzy Data Mining: Effect of Fuzzy Discretization", in: *Proceedings of the First IEEE International Conference on Data Mining, 2001*, pp. 241–248.
- [15] M. Setnes, R. Babuska, B. Verbruggen, "Rule-Based Modeling: Precision and Transparency", *IEEE Transactions on Systems, Man, and Cybernetics – Part C* 28 (1) (1998) 165–169.
- [16] Y. Jin, "Fuzzy Modeling of High-Dimensional Systems: Complexity Reduction and Interpretability Improvement", *IEEE Transactions on Fuzzy Systems* 8 (2) (2000) 212–221.
- [17] D.E. Goldberg, "Genetic Algorithms in Search Optimization and Machine Learning", Addison Wesley, Reading, MA, 1989.
- [18] M. Sugeno, T. Yasukawa, "A Fuzzy-Logic-Based Approach to Qualitative Modeling", *IEEE Transactions on Fuzzy Systems* 1 (1) (1993) 7–31.
- [19] L.A. Zadeh, "A Fuzzy-Set-Theoretic Interpretation of Linguistic Hedges", *Journal of Cybernetics* 2 (1972) 4–34.
- [20] J. Yen, L. Wang, "Application of Statistical Information Criteria for Optimal Fuzzy Model Construction", *IEEE Transactions on Fuzzy Systems* 6 (3) (1998) 362–372.
- [21] J. Yen, L. Wang, "Simplifying Fuzzy Rule-Based Models Using Orthogonal Transformation Methods", *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics* 29 (1) (1999) 13–24.
- [22] L. Wang, J. Yen, "Extracting Fuzzy Rules for System Modeling Using a Hybrid of Genetic Algorithms and Kalman filter", *Fuzzy Sets and Systems* 101 (1999) 353–362.
- [23] M. Setnes, H. Roubos, "GA-Fuzzy Modeling and Classification: Complexity and Performance", *IEEE Transactions on Fuzzy Systems* 8 (5) (2000) 509–522.
- [24] D. Agarwal, S.P. Tripathi, J.B. Singh., "TrFRA: A Trust Based Fuzzy Regression Analysis", Published in *IRECOS Vol.5 N.6 November 2010*, pp.668-670.
- [25] D. Agarwal, S.P. Tripathi, J.B. Singh, "Web Personalization of Indian e-Commerce Websites using Classification Methodologies", Published in *IJSCI Volume: 7 Issue: 6, November 2010*, pp. 18-21.

Implementation of Genetic Algorithm for Automatic Test Pattern Generation

MrsRachna singh, Dr.Arvind Rajawat

Abstract: This paper presents, genetic based algorithm for random test pattern generation .Genetic algorithm solve many search and optimization problem, effectively. In genetic algorithm, the optimized test vector is generated, which enhances the fault coverage and improve the global search .As, a result a new random-based test pattern generation technique based on GA was presented. Experiment results showed that the genetic algorithm improved the ability of global search and increases the fault coverage. This algorithm improves the test size with a factor of about 25% in comparison with other approaches for ATPG.

Keywords: Automatic Test Pattern Generation (ATPG), Genetic Algorithm (GA), Design Validation, Simulation Based Approaches, Evolutionary Algorithm.

1 INTRODUCTION:

In present scenario, digital systems are extremely intricate and increasing in complexity, which are required for use in widening range of domestic and industrial application. So to ensure reliability of these digital circuits, it is necessary to test their performance to identify any defects prior to using them in a fully operational environment.

These circuits are tested by test vectors. The test vectors are generated by efficient automatic test pattern generator (ATPG).The generation of test pattern with high fault coverage rate is a very expensive process for large circuits. An efficient ATPG tool reduces the test pattern generation time and cost, beside the high fault coverage rate.

There are many approaches for ATPG,like deterministic approach, simulators etc.The aim of this technique should be both to reduce execution time and to improve fault coverage.

Genetic algorithm described by Goldberg[13] is specially suited to solve large scale combination optimization problem.GA have been successfully applied in different areas of VLSI design

,especially in test branches such as test pattern generation [5].

2 MATERIAL AND METHODS:

2.1 Automatic Test Pattern Generation (ATPG):

ATPG is a method used to find an input (or test vectors) sequence that, when applied to a digital circuit, enables tester to distinguish between the correct behavior and the faulty circuit behavior caused by defects. The efficiency of this method is measured by using fault coverage, computational time and the length of test set.

Breuer [8] uses a fault simulator to evaluate sets of random vectors and to select the best vector to apply in each time frame. Weighted random pattern are interfaced with fault simulators [15] and high coverage is obtained for combinational circuits. The test generator [16] is built around fault simulator; the next candidate vector is generated using the hamming distance with the previous test vector.

2.2 Genetic Approach for Test PatternGeneration:

In past, test generation using deterministic & fault oriented algorithm is highly complex and time consuming new approaches are needed to augment the existing techniques, to reduce execution time and to improve fault coverage.GA was first used for simulation based test generation in [17].Several

Dr.ArvindRajawat is an Associate Professor ,MANIT ,
BHOPAL,India email:arvind.rajawat@gmail.com

Mrs.Rachna Singh is an Assistant Professor ,SISTEC-E ,
BHOPAL,Indiaemail:ahana10jan@gmail.com

IJSER © 2012
<http://www.ijser.org>

approaches to test generation have been proposed in [4], [9]. In reference [4], [9] the fitness evaluation and population scoring is low cost and only based on the fault coverage of each test vector. The disadvantage of the technique is that if a dropping fault simulation is used, experimentally after almost 10 generation, the generated vectors stop detecting remaining faults. This method has resulted in better final test set, but it is very expensive. A new operator is used in [4], in which, after each generation, the best vector in population is put on the final test set and then rescored with a new decreasing fitness.

2.2.1 Introduction:

Genetic algorithms are a part of evolutionary computing, which is a rapidly growing area of artificial intelligence. As you can guess, genetic algorithms are inspired by Darwin's theory about evolution. Simply said, solution to a problem solved by genetic algorithms is evolved. As a result, a new random based test pattern generation technique based on GA is presented. Experimental result show that this algorithm increases fault detection rate while test size decreases.

This is repeated until some condition (for example number of populations or improvement of the best solution) is satisfied.

2.2.2 Outline of the Basic Genetic Algorithm

1. **[Start]** Generate random population of n chromosomes (suitable solutions for the problem)
2. **[Fitness]** Evaluate the fitness $f(x)$ of each chromosome x in the population **[New population]** Create a new population by repeating following steps until the new population is complete
 1. **[Selection]** Select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to be selected)

2. **[Crossover]** With a crossover probability cross over the parents to form a new offspring (children). If no crossover was performed, offspring is an exact copy of parents.
3. **[Mutation]** With a mutation probability mutate new offspring at each locus (position in chromosome).
4. **[Accepting]** Place new offspring in a new population
3. **[Replace]** Use new generated population for a further run of algorithm

3 USING GENETIC ALGORITHM IN ATPG:

In this paper, a genetic algorithm approach to ATPG is used. The set of solutions called population is the test vectors. The purpose of this algorithm is to finding optimal solution with high convergence speed. A random population of n chromosome is generated and fitness of each chromosome in the population is evaluated. New population is created by repeating selection, crossover, mutation and acceptance. The test vector's generated by this algorithm cover's maximum number of faults in the VLSI circuits. Results show both a reduction in test sizes and an improvement in fault coverage compared with other results for combinational benchmark circuits.

The pseudo code of genetic algorithm for ATPG is shown in fig1. The description of each step of the algorithm is as follows:-

3.1 Generation of the Initial Antibodies:

In this step, the antibodies (test vectors) are created randomly on feasible space. Population size should be large enough in order to ensure adequate diversity; however, it is a trade off between getting higher convergence rate with larger search space and less genetic operation time. Population size in algorithm of Ivask, 1998 is constant value for all circuits. Experiments have proven that required population size increases with increasing test vector length. In this paper we used a more exact population size as shown in table 1

TABLE 1
POPULATION SIZE FOR DIFFERENT VECTOR
LENGTH

Population Size	Vector Length
8	<4
16	4-16
16	17-49
24	50-63
24	64-99
32	>99

3.2 Calculating the Fitness:

The fitness function provides a quantification of the quality of the chromosome.

It is the fitness of the chromosome that determines whether the chromosome will be selected to produce offspring and quantifies its chance for survival among the other chromosome in the population to the next generation. The fitness function is problem

Fig1. Genetic Algorithm Flow Chart

specific. In this paper fault simulation with fault is with fault dropping is used in order to evaluate the test vectors. The score given to each individual is equal to the number of fault it detects, the fitness function for given test vector is calculated by eq(1) given below:

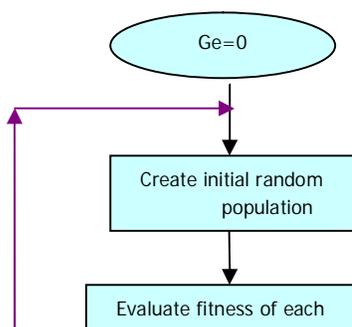
$$F(x) = \frac{\text{No of detected faults by test vector}}{\text{Total no of faults}}$$

3.3 Creating a New population:

New population is created by repeating the following steps until the new population is complete:

3.3.1 Selection:

The genetic algorithm uses selection operator to simulate natural evolution. In GA, individual with high fitness is inherited to the next generation with greater probability. Usually, chromosomes with high fitness are selected for crossover to converge faster to best solution. Chromosomes with high fitness should not be selected for mutation to prevent the danger of diverting from good solutions in the search space. Therefore, chromosomes with low fitness are usually selected for mutation. All selection methods are based



on the fitness of chromosomes. The disadvantage of selecting chromosomes with high fitness is the probability of less diversity in the search space. Therefore, chromosome with low fitness is usually selected for mutation. All Selection methods are based on the fitness of the chromosome.

In this paper Roulette wheel selection is used to select the individuals.

Roulette Wheel Selection:

Parents are selected according to their fitness. The better the chromosomes are, the more chances to be selected they have. Imagine a roulette wheel where are placed all chromosomes in the population, every chromosome has its place big accordingly to its fitness function, like on the following picture.

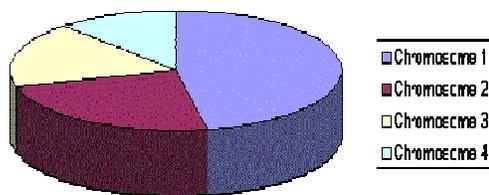


Fig 2

Then a marble is thrown there and selects the chromosome. Chromosome with bigger fitness will be selected more times.

This can be simulated by following algorithm.

1. **[Sum]** Calculate sum of all chromosome fitnesses in population - sum S .
2. **[Select]** Generate random number from interval $(0, S) - r$.
3. **[Loop]** Go through the population and sum fitnesses from 0 - sum s . When the sum s is greater than r , stop and return the chromosome where you are

Of course, step 1 is performed only once for each population.

3.3.2 Crossover

Crossover is the key to genetic algorithm, power that is to exchange corresponding genetic properties from

the two parents, to allow useful genes on different parents to combine in their offspring. Most common crossover types are one-point, two-point, uniform crossover. In this paper, as shown in fig (3), two-point crossover is used

3.3.3 Mutation:

After a crossover is performed, mutation takes place. This is to prevent falling all solutions in population into a local optimum of solved problem. Mutation changes randomly the new offspring. For binary encoding we can switch a few randomly chosen bits from 1 to 0 or from 0 to 1. It says how often will be parts of chromosome mutated. If there is no mutation, offspring is taken after crossover (or copy) without any change. If mutation is performed, part of chromosome is changed. If mutation probability is 100%, whole chromosome is changed, if it is 0%, nothing is changed.

Mutation is made to prevent falling GA into local extreme, but it should not occur very often, because then GA will in fact change to random search.

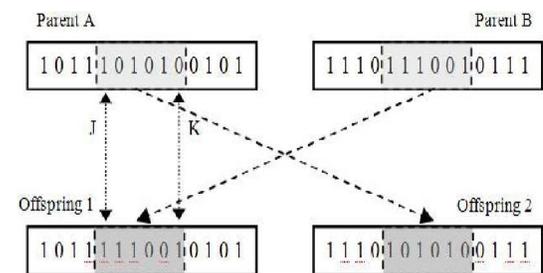


Fig3. Two-point crossover

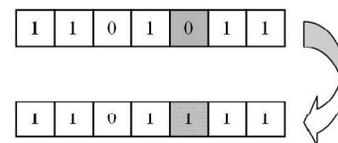


Fig4. Mutation in binary string

4. EXPERIMENT AND RESULTS:

The experiment is carried on the combinational circuits. These circuits are simulated for fault coverage ie, the number of faults covered by applying a test pattern to the circuit.

The Genetic algorithm is simulated using c language .The experiments are carried out on ISCAS'85 combinational benchmark circuits & other combinational circuits.

1. Experimental result shows that the test pattern size has been reduced up to 25% in comparison with other methods like verifault simulators as shown in table2.
2. Similarly the results show that the fault coverage is increased in genetic algorithm in comparison with verifault simulator as shown in table 3.

5. CONCLUSION:

This paper uses genetic algorithm for automatic test pattern generation in VLSI circuits. ATPG tools can reduce the amount of effort and cost of test generation.

TABLE 2
REDUCTION IN TEST SIZE USING IGA

Our experimental results showed that genetic algorithm based method are more efficient in test generation in comparison with other approaches for ATPG.

The experimental results shows an

- Average 25% reduction in test size
- Increase in fault coverage.

TABLE 3
RESULT OF FAULT SIMULATION ON
COMBINATIONAL CIRCUITS

Circuit	Total no of Detected fault by Verifault Simulator	Total no of Detected fault by GA	% Improvement in fault detected
C17	19	22	4.9%
1 bit adder	22	24	8.3%
2 to 1 MUX	13	14	7.2%

REFERENCES

- [1].Azimipour, M. and M. Eshghi, Parallel Circular- Scan Architecture. Asian Network for Scientific Information (ANSI), Journal of applied sciences, pp: 1-8,2008.
- [2].Azimipour, M. Eshghi and A. Khademzadeh,. A Modification to Circular-Scan Architecture to

Circuit	No of Nodes		Test size		Improve ment
	inp ut	outp ut	Verifault fault simulator(Rad,M.A,S.M. Eshgh,2007)	Genetic Algo	
c 17	5	2	4	3	25%
1 bit adder	3	2	4	3	25%
2 to 1 MUX	3	1	5	4	20%

improve test data compression. *IEEE CS Proc. Of 15th Int. conf. on advance computing and communication*,pp: 27-33,2007.

[3].Al-Yamani, A., Erik Chmelar and Mikhail Grinchuk.Segmented Addressable Scan Architecture. *In Proc. of VLSI Test Symposium*, pp: 405-411,2005.

[4]. Rad, M.A. and S.M. Eshgh. A heuristic Algorithm with a revocation based GA for test Pattern generation of VLSI circuits. *IEEE Proc. Of Int Conf. On Electrical Engineering*, pp: 1-5,2007.

[5]. Li, Y., Y. Dai and X. Ma. A heuristic immune-genetic algorithm for multimodal function optimization. *In Proceeding of CIMCA/IAWTIC*

conference, pp: 36-40, 2005.

[6].Chattopadhyay, S. and N. Choudhary. Genetic algorithm based approach for low power combinational circuit testing. *In proceeding of the IEEE 16th Int. Conference on VLSI Design*, pp: 552-557, 2003.

[7] Jiaxin, Y., C. Baichao, T. Cuihua and S. Yu. The research of inverter's control based on immune genetic algorithm. *In Proc. Of IEEE Conference on Industrial Electronics and Applications*, pp: 1-6, 2006.

[8].Breuer, A. and A.D. Friedman. Diagnosis and reliable design of digital systems: *Computer Science Press Inc.* , 1997

[9].Arslan, T. and M.J., O'Dare. A genetic algorithm for multiple fault model test generation. *In Proc. of Second International Conference on Genetic Algorithms in Engineering Systems. Innovations and Applications*, pp: 462-466 , 1997

[10].Al-Yamani, A., Erik Chmelar and Mikhail Grinchuk. Segment Addressable Scan Architecture. *In Proc. of VLSI Test Symposium*, pp: 405-411, 2005

[11].Essentials Of Electronic Testing For Digital, Memory And Mixed- Signal VLSI Circuits , by Michael L. Bushnell (*Rutgers University*)
Vishwani D. Agrawal (*Bell Labs, Lucent Technologies*), Kluwer Academic Publishers, 2002.

[12] .Introduction to genetic algorithm –tutorial, by Mark obitko,
www.obotiko.com/tutorial/genetic-algorithm

[13].Goldberg,D.E, "genetic algorithms in search,optimization and Machine Learning.Reading,MA:Addison-Wesley.1989.

[14].Jiao ,L. and L. Wang,2000., "A novel genetic algorithm based on immunity".PERLINK<http://www.informatik.uni-trier.de/ley/db/journals/tsmc/tsmca30.html> \1" JiaoW00" IEEE Transactions on systems,Man and Cybernetics,Parta 30(5):552-561,2000.

[15].Schnurmann,H.D,E.Lindbloom and R.G.Carpenter," The weighted random test –pattern generator. *IEEETrans.Comput.*, VOL. c-24,pp:695-700,1975.

[16].Snethen,T.J., "Simulator Oriented fault test generator" *Inproc.Of design Automation Conf.*,pp:88-93,1977.

[17].Seshu,S. and D.N.Freeman," The diagnosis osasynchronous sequential switching systems,*IRE Trans on electronics Computing*,vol.EC-11,pp:459-465,1962

Some New Trigonometric, Hyperbolic and Exponential Measures of Fuzzy Entropy and Fuzzy Directed Divergence.

Jha P., MishraVikas Kumar

Abstract: New Trigonometric, Hyperbolic and Exponential Measures of Fuzzy Entropy and Fuzzy Directed Divergence are obtained and some particular cases have been discussed.

Index Terms: Fuzzy Entropy, Fuzzy Directed Divergence, Measures of Fuzzy Information.



1. Introduction: Uncertainty and fuzziness are the basic nature of human thinking and of many real world objectives. Fuzziness is found in our decision, in our language and in the way we process information. The main use of information is to remove uncertainty and fuzziness. In fact, we measure information supplied by the amount of probabilistic uncertainty removed in an experiment and the measure of uncertainty removed is also called as a measure of information while measure of fuzziness is the measure of vagueness and ambiguity of uncertainties. Shannon [2] used “entropy” to measure uncertain degree of the randomness in a probability distribution. Let X is a discrete random variable with probability distribution $P = (p_1, p_2, \dots, p_n)$ in an experiment. The information contained in this experiment is given by

$$H(P) = - \sum_{i=1}^n p_i \log p_i \quad (1)$$

Which is well known Shannon entropy.

The concept of entropy has been widely used in different areas, e.g. communication theory, statistical mechanics, finance, pattern recognition, and neural network etc. Fuzzy set theory developed by Lofti A. Zadeh [8] has found wide applications in many areas of science and technology, e.g. clustering, image processing, decision making etc. because of its capability to model non-statistical imprecision or vague concepts.

It may be recalled that a fuzzy subset A in U (universe of discourse) is characterized by a membership function $\mu_A: U \rightarrow [0,1]$ which represents the grade of membership of $x \in U$ in A as follows

$$\mu_A(x) = 0 \text{ if } x \text{ does not belongs to } A, \\ \text{and there is no uncertainty}$$

$= 1$ if x belongs to A and there is no uncertainty

$= 0.5$ if maximum uncertainty

In fact $\mu_A(x)$ associates with each $x \in U$ a grade of membership in the set A . When $\mu_A(x)$ is valued in $\{0,1\}$ it is the characteristic function of a crisp (i.e. nonfuzzy) set. Since $\mu_A(x)$ and $1 - \mu_A(x)$ gives the same degree of fuzziness, therefore, corresponding to the entropy due to Shannon [2], De Luca and Termini [1] suggested the following measure of fuzzy entropy:

$$H(A) = - \left[\sum_{i=1}^n \mu_A(x_i) \log \mu_A(x_i) + \sum_{i=1}^n (1 - \mu_A(x_i)) \log (1 - \mu_A(x_i)) \right] \quad (2)$$

De Luca and Termini introduced a set of properties and these properties are widely accepted as a criterion for defining any new fuzzy entropy. In fuzzy set theory, the entropy is a measure of fuzziness which expresses the amount of average ambiguity/difficulty in making a decision whether an element

belongs to a set or not. So, a measure of average fuzziness in a fuzzy set should have at least the following properties to be valid fuzzy entropy:

- i) $H(A) = 0$ when $\mu_A(x_i) = 0$ or 1 .
- ii) $H(A)$ increases as $\mu_A(x_i)$ increases from 0 to 0.5.
- iii) $H(A)$ decreases as $\mu_A(x_i)$ increases from 0.5 to 1.
- iv) $H(A) = H(\bar{A})$, i.e. $\mu_A(x_i) = 1 - \mu_A(x_i)$
- v) $H(A)$ is a concave function of $\mu_A(x_i)$.

Kullback and Leibler [7] obtained the measure of directed divergence of probability distribution $P = (p_1, p_2, \dots, p_n)$ from the probability distribution $Q = (q_1, q_2, \dots, q_n)$ as

$$D(P:Q) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i} \quad (3)$$

Let A and B be two standard fuzzy sets with same supporting points x_1, x_2, \dots, x_n and with fuzzy vectors $\mu_A(x_1), \mu_A(x_2), \dots, \mu_A(x_n)$ and $\mu_B(x_1), \mu_B(x_2), \dots, \mu_B(x_n)$. The simplest measure of fuzzy directed divergence as suggested by Bhandari and Pal (1993), is

$$D(A:B) = \sum_{i=1}^n \mu_A(x_i) \log \frac{\mu_A(x_i)}{\mu_B(x_i)} + \sum_{i=1}^n (1 - \mu_A(x_i)) \log \frac{(1 - \mu_A(x_i))}{(1 - \mu_B(x_i))} \quad (4)$$

satisfying the conditions:

- i) $D(A:B) \geq 0$
- ii) $D(A:B) = 0$ iff $A = B$
- iii) $D(A:B) = D(B:A)$
- iv) $D(A:B)$ is a convex function of $\mu_A(x_i)$

later kapur [5],[6] introduced a number of trigonometric hyperbolic and exponential measures of fuzzy entropy and fuzzy directed divergence. In section 2 and 3 we introduce some new trigonometric, hyperbolic and exponential measures of fuzzy entropy and measures of fuzzy directed divergence.

2. New Measures of Fuzzy Entropy

2.1 Trigonometric Measure of Fuzzy Entropy

Consider the function $\sin \pi x$ where $0 \leq x \leq 1$, is a convex function which gives us

$$H_1(A) = \sum_{i=1}^n \sin(\pi \mu_A(x_i)) + \sum_{i=1}^n \sin(\pi(1 - \mu_A(x_i))) \quad (5)$$

is a new measure of fuzzy entropy.

in particular for $\beta \leq \pi$

$$H_2(A) = \sum_{i=1}^n \sin(\beta \mu_A(x_i)) + \sum_{i=1}^n \sin(\beta(1 - \mu_A(x_i))) - \sin \beta \quad (6)$$

is also a new measure of fuzzy entropy.

(5) is a special case of (6) when $\beta = \pi$.

Another special case of (6) arises when $\beta = \frac{\pi}{2}$ we get

$$H_3(A) = \sum_{i=1}^n \sin\left(\frac{\pi}{2} \mu_A(x_i)\right) + \sum_{i=1}^n \sin\left(\frac{\pi}{2} (1 - \mu_A(x_i))\right) - 1 \quad (7)$$

Another trigonometric measure of fuzzy entropy is

$$H_4(A) = \sum_{i=1}^n \sin(\beta \mu_A(x_i) + \alpha) + \sum_{i=1}^n \sin(\beta(1 - \mu_A(x_i)) + \alpha) - \sin(\alpha + \beta) \quad (8)$$

(8) reduces to (6) when $\alpha = 0$.

(8) reduces to (7) when $\alpha = 0, \beta = \frac{\pi}{2}$.

(8) reduces to (5) when $\alpha = 0, \beta = \pi$.

(8) is a 2-parameter measure of fuzzy entropy.

If we put $\alpha = \frac{\pi}{2}$ we get

$$H_5(A) = \sum_{i=1}^n \cos(\beta \mu_A(x_i)) + \sum_{i=1}^n \cos(\beta(1 - \mu_A(x_i))) - \cos \beta \quad (9)$$

is a new measure of fuzzy entropy. Clearly above given measures of fuzzy entropy are satisfying all the properties which are given in section 1. So these are valid measures of fuzzy entropy.

2.1 Hyperbolic Measure of Fuzzy Entropy

$\sinh x, \cosh x, \tanh x$ where $0 \leq x \leq 1$ are all convex functions and gives us following valid measures of fuzzy entropy

$$H_6(A) = \sinh\beta - \sum_{i=1}^n \sinh(\beta\mu_A(x_i)) - \sum_{i=1}^n \sinh\beta(1 - \mu_A(x_i)) \quad (10)$$

$$H_7(A) = \cosh\beta - \sum_{i=1}^n \cosh(\beta\mu_A(x_i)) - \sum_{i=1}^n \cosh\beta(1 - \mu_A(x_i)) \quad (11)$$

$$H_8(A) = \tanh\beta - \sum_{i=1}^n \tanh(\beta\mu_A(x_i)) - \sum_{i=1}^n \tanh\beta(1 - \mu_A(x_i)) \quad (12)$$

Since $x^m \sinh x, x^m \cosh x, x^m \tanh x$ are also convex functions for $m \geq 1$, we get the following additional measures of fuzzy entropy.

$$H_9(A) = \sinh\beta - \sum_{i=1}^n \mu_A^m(x_i) \sinh(\beta\mu_A(x_i)) - \sum_{i=1}^n (1 - \mu_A(x_i))^m \sinh\beta(1 - \mu_A(x_i)) \quad (13)$$

$$H_{10}(A) = \cosh\beta - \sum_{i=1}^n \mu_A^m(x_i) \cosh(\beta\mu_A(x_i)) - \sum_{i=1}^n (1 - \mu_A(x_i))^m \cosh\beta(1 - \mu_A(x_i)) \quad (14)$$

$$H_{11}(A) = \tanh\beta - \sum_{i=1}^n \mu_A^m(x_i) \tanh(\beta\mu_A(x_i)) - \sum_{i=1}^n (1 - \mu_A(x_i))^m \tanh\beta(1 - \mu_A(x_i)) \quad (15)$$

2.2 Exponential Measures of Fuzzy Entropy

Since $x^m e^{ax}$ is a convex function when $m \geq 1, x > 0$ we get the measure of fuzzy entropy

$$H_{12}(A) = e^a - \sum_{i=1}^n \mu_A^m(x_i) e^{a\mu_A(x_i)} - \sum_{i=1}^n (1 - \mu_A(x_i))^m e^{a(1 - \mu_A(x_i))} \quad (16)$$

3. New Measures of Fuzzy Directed Divergence

3.1 New Hyperbolic Measures of Fuzzy Directed Divergence

Using the convexity of $\sinh x, \cosh x, \tanh x$ we get the following measures of hyperbolic fuzzy

directed divergence.

$$D_1(A:B) = \sum_{i=1}^n \mu_B(x_i) \sinh\left(\beta \frac{\mu_A(x_i)}{\mu_B(x_i)}\right) + \sum_{i=1}^n (1 - \mu_B(x_i)) \sinh\left(\beta \frac{(1 - \mu_A(x_i))}{(1 - \mu_B(x_i))}\right) - \mu_B(x_i) \sinh\beta - (1 - \mu_B(x_i)) \sinh\beta \quad (17)$$

$$D_2(A:B) = \sum_{i=1}^n \mu_B(x_i) \cosh\left(\beta \frac{\mu_A(x_i)}{\mu_B(x_i)}\right) + \sum_{i=1}^n (1 - \mu_B(x_i)) \cosh\left(\beta \frac{(1 - \mu_A(x_i))}{(1 - \mu_B(x_i))}\right) - \mu_B(x_i) \cosh\beta - (1 - \mu_B(x_i)) \cosh\beta \quad (18)$$

$$D_3(A:B) = \sum_{i=1}^n \mu_B(x_i) \tanh\left(\beta \frac{\mu_A(x_i)}{\mu_B(x_i)}\right) + \sum_{i=1}^n (1 - \mu_B(x_i)) \tanh\left(\beta \frac{(1 - \mu_A(x_i))}{(1 - \mu_B(x_i))}\right) - \mu_B(x_i) \tanh\beta - (1 - \mu_B(x_i)) \tanh\beta \quad (19)$$

Again since $x^m \sinh x, x^m \cosh x, x^m \tanh x$ are also convex functions for $m \geq 1$, we get the following more general hyperbolic measures of fuzzy directed divergence.

$$D_4(A: B) = \sum_{i=1}^n \mu_A^m(x_i) \mu_B^{1-m}(x_i) \sinh \left(\beta \frac{\mu_A(x_i)}{\mu_B(x_i)} \right) + \sum_{i=1}^n (1 - \mu_A(x_i))^m (1 - \mu_B(x_i))^m$$

$$\sinh \left(\beta \frac{(1 - \mu_A(x_i))}{(1 - \mu_B(x_i))} \right)$$

$$- \mu_B(x_i) \sinh \beta - (1 - \mu_B(x_i)) \sinh \beta \quad (20)$$

$$D_5(A: B) = \sum_{i=1}^n \mu_A^m(x_i) \mu_B^{1-m}(x_i) \cosh \left(\beta \frac{\mu_A(x_i)}{\mu_B(x_i)} \right) + \sum_{i=1}^n (1 - \mu_A(x_i))^m (1 - \mu_B(x_i))^m$$

$$\cosh \left(\beta \frac{(1 - \mu_A(x_i))}{(1 - \mu_B(x_i))} \right)$$

$$- \mu_B(x_i) \cosh \beta - (1 - \mu_B(x_i)) \cosh \beta \quad (21)$$

$$D_6(A: B) = \sum_{i=1}^n \mu_A^m(x_i) \mu_B^{1-m}(x_i) \tanh \left(\beta \frac{\mu_A(x_i)}{\mu_B(x_i)} \right) + \sum_{i=1}^n (1 - \mu_A(x_i))^m (1 - \mu_B(x_i))^m$$

$$\tanh \left(\beta \frac{(1 - \mu_A(x_i))}{(1 - \mu_B(x_i))} \right)$$

$$- \mu_B(x_i) \tanh \beta - (1 - \mu_B(x_i)) \tanh \beta \quad (22)$$

3.2 New Exponential Measures of Fuzzy Directed Divergence

Since $x^m e^{ax}$ is a convex function when $m \geq 1, x > 0$ we get the following measures of fuzzy directed divergence

$$D_7(A: B) = \sum_{i=1}^n \mu_A^m(x_i) \mu_B^{1-m}(x_i) e^{a \left(\frac{\mu_A(x_i)}{\mu_B(x_i)} \right)} + \sum_{i=1}^n (1 - \mu_A(x_i))^m (1 - \mu_B(x_i))^{1-m} e^{a \left(\frac{1 - \mu_A(x_i)}{1 - \mu_B(x_i)} \right)} - e^a \quad (23)$$

Special case for $m=0$ and $m=1$ are

$$D_8(A: B) = \sum_{i=1}^n \mu_B^m(x_i) e^{a \left(\frac{\mu_A(x_i)}{\mu_B(x_i)} \right)} + \sum_{i=1}^n (1 - \mu_B(x_i)) e^{a \left(\frac{1 - \mu_A(x_i)}{1 - \mu_B(x_i)} \right)} - e^a \quad (24)$$

$$D_9(A: B) = \sum_{i=1}^n \mu_A^m(x_i) e^{a \left(\frac{\mu_A(x_i)}{\mu_B(x_i)} \right)} + \sum_{i=1}^n (1 - \mu_A(x_i)) e^{a \left(\frac{1 - \mu_A(x_i)}{1 - \mu_B(x_i)} \right)} - e^a \quad (25)$$

4. Conclusion

In section 2 and 3 by using the convexity of some trigonometric, hyperbolic and exponential function and satisfying the conditions of fuzzy entropy and fuzzy directed divergence we get some new trigonometric, hyperbolic and exponential measures of fuzzy entropy and fuzzy directed divergence.

Jha P. Department of Mathematics, Govt Chattisgarh P.G. College,
Raipur, Chattisgarh(India)

Mishra Vikas Kumar Department of Mathematics, Rungta College
of Engineering and Technology, Raipur, Chattisgarh(India)

5. References

[1] A. De Luca and S. Termini, *A Definition of a Non-probabilistic Entropy in the Setting of fuzzy sets theory*, Information and Control, 20, 301-312, 1972.

[2] C.E.Shannon (1948). "A Mathematical Theory of Communication". Bell. System Tech. Journal Vol. 27, pp. 379-423,623-659.

[3] D. Bhandari and N. R. Pal, *Some new information measures for fuzzy sets*, Information Science, 67, 204 - 228, 1993.

[4] F.M.Reza (1948&1949). "An introduction to information theory". Mc. Graw-Hill, New-York.

[5] J.N.Kapur. Some New Measures of Directed Divergence. Willey Eastern Limited.

[6] J.N.Kapur. Trigonometrical Hyperbolic and Exponential Measure of Fuzzy Entropy and Fuzzy Directed Divergence. MSTs

[7] S. Kullback and R.A. Leibler, *On Information and Sufficiency*, Annals of Mathematical Statistics, 22, 79 - 86, 1951.

[8] L. A. Zadeh, *Fuzzy Sets*, Information and Control, 8, 338 - 353, 1965.

Residents Awareness Towards Landslide Disaster In Penang

Habibah Lateh, Vijaya Govindasamy

Abstract— This study aims to identify the resident awareness of landslides disaster. The objective of the study is to identify the level of resident awareness of landslides disaster in Penang. A total of 240 people representing ten of each of the six districts in the state of Penang in the sample by simple random sampling method. Data collected is analyzed using the Statistical Package For Social Sciences (SPSS) using percentages and frequencies. The results showed that the level of resident awareness of landslides is moderate. This means that the government and political parties need to increase resident awareness of landslides disaster. In addition, the knowledge resident of landslides is moderate too. So still plenty of room for improvement to ensure a landslide awareness program carried out properly and transparently.

Keywords— Minimum 7 disaster, environmental problem, knowledge, residents, awareness, Penang, landslide.

◆

1 INTRODUCTION

Highlands and hillsides are one of the natural resources of the state of Penang. It is part of the environment and is very valuable because besides that it is contributing to the economy and also keeps a treasure of flora and fauna of Penang. People staying near the hills are often exposed to the activities of exploration of hills and mountains with a slope greater than 25 degrees and above. This phenomenon becomes more complex when there are more and more of a greedy and irresponsible people baring the hilly areas. Their greed is not only invites disaster to the resident, property and the environment but also cause severe losses to the government. Rapid development along the hillside have been made to meet the current demand. Each development undertaken is not denied, would lead to certain implications. Similarly, the development carried out in the hills would have implications on the environment and also the safe of the residents staying on the hills. Among the implications arising from the exploration on the hillside causing excessive environmental problems such as soil erosion and landslides, water pollution, rising temperatures, the original placement disruption, loss of habitat for flora and fauna, sedimentation and flooding phenomena. Landslide awareness consideration is important in assessing the quality of life since the environment impact directly on the well-being of the residents. Many of the guidelines and the Act has been prepared to address the landslide, but landslides continue to happen in this state. Although there is a supporting strategy for landslide, such as monitoring landslide, landslide research, dissemination of landslide information, landslide training and also inter-agency cooperation and coordination, but there is no other on-going landslide programs strongly supporting strategy for landslide awareness and knowledge or landslide education in schools or public, in addressing landslide occurrences.

Landslide or landslip is a geological phenomenon which includes a wide range of ground movement, such as rock falls, deep failure of slopes and shallow debris flows, which can occur in offshore, coastal and onshore environments. Although the action of gravity is the primary driving force for a landslide to occur, there are other contributing factors affecting the original slope stability. By human exploitation along

the hill for various purposes, resulting in soil erosion as well as causing much loss to the residents who live nearby. This is a normal event that occurs in the hilly areas, but not to be taken seriously because of lack of response from the majority party and its residents. Furthermore hills are mostly owned by individuals and difficult to prevent them from developing. Because it gets the attention that is not so serious, so landslides are not taken seriously by certain parties, including residents who are brave enough to build their houses on the hill. For example, landslide that occurred on Taman Bukit Mewah, Bukit Antarabangsa Luxury Hill Park, Hill International, Ampang, Selangor on December 6, 2008 is due to the negligence of greedy people who cannot control development on hills which claimed many lives and destroyed property worth. Frequent landslides involving loss of life, but the action taken only after the landslide and it continues to forgotten or neglected. No matter what, landslides are caused by lack of awareness and knowledge of various parties including its residents. They are less aware of the disasters set up home in a dangerous slopes. This recurrence lead to deterioration of the environment and affect the state of Penang source of income, which is an important tourist center in Malaysia. This is due to the ignorance of various parties who are not aware and not knowledgeable of landslide disaster. Many green plants have been destroyed to open the forest for the construction. Animals become extinct due to human activities which do not have compassion on the destruction of this beautiful nature. This activity invites other disasters such as flash floods, sedimentation, and so forth. All levels of society should be aware and must have a high knowledge in order to maintain a natural environment for future generations. Landslides, flash floods and so on have contributed to severe environmental degradation in both rural and urban areas of Penang, and threaten human life and other wildlife on a large scale. Less focus on the preservation of nature and more focused in making a profit lead to catastrophic landslides. Landslide problems underestimated by most parties as well as residents. A major challenge for all parties involved in addressing the landslide is to translate policy intervention landslides in a landslide awareness programs, projects and other initiatives to

help advance the aim of this study to increase resident awareness of the dangers of landslides. Hence the importance of landslide disaster awareness programs to reduce the occurrence of landslides in the state with the support of all parties such as government, private or resident in dealing with this problem.

2 LITERATURE REVIEW

Environment includes all living and non-living objects. We live in the environment and use the land to meet our needs. Development also means meeting the needs of the people. While meeting the ever-growing needs, we put pressure on the environment. When the pressure exceeds the carrying capacity of the environment to repair or replace itself, it creates a serious problem of environmental degradation such as landslide, soil erosion, flood and etc. If we use any environmental resource such as hills and mountains beyond its limit of replacement, we may lose it forever. Therefore, there is a need to create 'awareness' about hills and mountains protection. While efforts are being made at the national and international level to protect our environment, it is also the responsibility of every residents to use our environmental resources with care and protect them from degradation. This reciprocity between self-produced action and environmental events lies at the heart of the development of a sense of competence [1]. Environment constitutes a very important part of our life. To understand life without studying the impact of environment is simply impossible. The need to protect environment can be ignored only at our peril. We use environmental resources in our daily life. These resources are renewable and non-renewable. We have to be more cautious in consuming non-renewable resources like hills and mountains, which are prone to natural disasters. All human activities have an impact on exploration of the hills. But lately, the human influence on the environment has increased with the increase in population and human needs. In the past two decades, landslide has attracted the attention of decision makers, scientists and even residents in many parts of the world. They are becoming increasingly conscious of this issue. People are now aware of the need to protect the natural environmental resources of soil or land and plant life that constitute the natural capital on which man depends on. Landslide issues began to be taken seriously because of the lack of solutions to the problem. If these issues are not taken seriously, the next generation may have a miserable life in facing the challenges of natural disasters. So, a better understanding of the issues among the residents of landslides and other parties involved directly or indirectly be enhanced and implemented properly in order to create a caring society on the environment as a whole. According to a central principle of ecological psychology developed by James and Eleanor Gibson [2], [3], [4] to be a living organism (to be animate) means to have agency that functions in the service of maintaining an awareness of the environment.

Penang island is a small state, so it is not surprising that developers have targeted the hill land for all sorts of development ranging from housing to transportation, industries and recreation. More than 50% of the Penang Island is made up of

steep topography above 60m [5]. As such, hill land is abundantly available on the island. Penang is one of the most rapid industrialisation, fast technological change, highest rate of urbanization and lead to greater demands for land. As such, it is anticipated that developers will cast their eyes on the remaining hill land on the island. Two districts from the northeast and southwest of Penang Island were selected which were situated at the prone area of landslide disasters. Therefore, it would be very interesting to examine whether the two districts residents are aware about the landslide disasters. So, the awareness among the residents are very important to prevent future landslides. Public Awareness Unit is responsible for training personnel in Public Work Departments (PWD) and creating public awareness [6]. Human consciousness always occurs in relation to the environment, and with the nature and man are not two separate things, but both should be shared. In a review by [7], important variables include biospheric values (care for plants, animals, and communities of living things) and the self-perceived ability to reduce threats to the environment. Awareness of landslides arising from the knowledge possessed by residents who live in areas prone to landslide hazards disaster. Residents awareness through various channels can be combined with technical knowledge taken from culture residents to raise awareness and confidence in living and empower them to act when faced with adversity.

Awareness provides a secure environment and can address the problems faced by residents in the event of landslides. Two examples of how residents knowledge can be used to heighten awareness and develop balanced environment of early precautions are, first to enjoined the spirit of planting greenery, and secondly, be exposed to incidents of a landslide disasters. These and other techniques are among the few indigenous early precautions which can serve as sign to prevent the incidence of landslides. They can be used to alert resident of an impending landslide disaster and to seek ways not to destroy the hilly areas reserves or increase the concrete buildings of maintaining green plants. On a regional scale, weather experts have already warned of severe landslides and floods after heavy rains, while Penang meteorologists predicted a wet season in the coming month. Landslide damage may also occur when hill lands are exhausted and no maintenance are made in their protection. At present there are several laws governing environment and landslide management issues, but there is no umbrella landslide legislation or a broader policy framework which deals with all landslide concerns. Landslide become one of the major issue in the state of Penang. The erosion of the soil surface and which makes soil heavier occur in the slopes due to the direct exposure of the soil surface to the rainfall. As we know, Penang is experiencing monsoon season, causing heavy rain throughout the year. Moreover, runoff of the surface proportionally increases and infiltrate the pore water pressure. As this flow is downward, this washes down the soil and causes the soil erosion. Finally, this type of erosion could cause major landslides.

All over the world, each year landslide disasters take a huge toll in deaths, property damage and economic loss. Disasters have always been part of the human experience [8]. Resident's perceptions of landslide disasters are influence by their

age, educational background and experiences. Research on residents landslide knowledge becomes vital if it is necessary to develop citizens who can take responsibility to protect and improve the environment. In this way, landslide awareness program or landslide education are underpinned with principles and attitudes that will benefit the residents. If the residents are aware of the landslide disasters for the long-term procedures faced by the them, they are more likely to take steps to reduce potential losses. While landslide research and landslide protection policy has ever since been dominated by a technical world view, the social and socio-economic aspects gained in importance in recent decades due to expansive and intensified hill land use, rising damage potentials in landslide areas and, thus, increasing conflicts between socio-economic on hill land use and landslide protection policy. While there are many ways that people form their attitudes [9],[11], two are relevant here, media and direct personal experience. So, the residents can be aware of landslide disaster by their personal experience, too. During the last years a shift in paradigms can be observed from a technical oriented landslide protection towards landslide awareness, which takes all societal advantages and disadvantages, or in economic terms, all benefits and costs of different landslide awareness strategies into account. Design deficiency can also cause landslides and this has been acknowledged by experts saying a review of the causes of landslides indicates that most of the landslides are man-made slopes and are mainly due to design deficiency [10]. [12] define stakeholders as individuals or groups who may receive a significant impact from the development of a project or program and the World Development Council for Sustainable Development (2003) is disclosed as a residents or institutions who are affected, or may be affected, by the activities of certain organizations. [13] cite Buttell's observation that the items used in his study "may tap a somewhat superficial awareness of environmental problems rather than a more meaningful commitment to solve such problems". This indicates that residents should be more positive for the environment before the disaster occurred and take steps too late.

Over recent decades, global problems relating landslide disasters have increased dramatically. It is caused by excessive hill land use. Lands scarcity on a global scale, deforestation, degradation of hill areas, soil erosion and loss of biodiversity, are some of the problems that have become a major concern. Although high rainfalls is the cause of the landslide, but the man-made factors will be the main cause of landslides. Most of the landslides in two monsoon seasons of Malaysia are induced by the high rainfalls and more than 80% of landslides were caused by man-made factors, mainly design and construction errors [10]. Destruction of forests and mountains have caused many environmental problems such as landslides, soil erosion and etc have reached such levels that have already resulted in serious environmental problems, as well as negative impact on the environment, and inevitably influencing prospects for long-term economic growth. With heavy rainfall throughout the year in many Asian countries, hill or mountain land have become prone to landslides, erosion, flash floods, lightning or rain, global warming due to rampant felling of green plants have taken a toll on the coun-

tries here. Now the landslide problems have an adverse effect upon quality of life. Disaster tragedies as well as landslide disasters due to extensive development at hill slopes given rise to the public concern in recent years. Recently, development of hill areas and hill slopes have increased risks of landslides which caused substantial loss of human life and damages to the properties and infrastructures. The most important lesson from experience is that prevention of disaster is better than trying to revenge the damage once it has occurred as stated in Agenda 21 of the United Nations Conference on Environment and Development 1992 [14].

Landscape no longer alien to residents in this state, especially in urban areas. An increasing number of private construction of luxury houses on hills, residential premises, hotels, recreational parks, resorts, tourism areas, other basic facilities began to emphasize the green environment and to illuminate the art of beautiful and interesting landscapes. [15] suggested that attitudes must be based on knowledge, or to restate this in terms of attitude theory, a general positive or negative orientation toward the environment must be based on beliefs. Knowledge and awareness of a person can change the environment to safe or unsafe. [16] found that those with more knowledge did have more favorable environmental attitudes and held themselves more responsible for pro-environmental acts. [17] showed that for a sample of Illinois residents those with more knowledge had more positive environmental attitudes. Younger people, well educated people, and politically liberal people tend to be more concerned about environmental quality than their older, less educated, and politically conservative counterparts. Attending an environmental training camp can improve the knowledge and awareness of landslides among the residents. In a field experiment [18] did a more convincing test of the hypothesis by showing that those who attended a camp which increased knowledge had both greater knowledge and a more liberal attitude toward forest management than a control group.

In this contribution it is argued that the current challenge in landslide awareness research consists in developing a better understanding of the interrelations and social dynamics of landslide knowledge, perception, preparedness, vulnerability, deep concern about the destruction of green plants on hilly areas, landslide management, landslide policy and etc, to take this into account in a modern design of landslide awareness programs. The relationship between landslide awareness and knowledge are two things that should be emphasized in addressing the problem of landslides in the state or across the country. The discusses of the shortcomings of the current approaches with a special focus on landslide awareness among the residents in this island and disregard for rapid development along the hill slopes until destroyed the natural environment of this state. In fact, the hill is an important component of tourism and cultural heritage of the state. If this state is engulfed by concrete buildings, then this will cause 'heat island' solely in the name of 'development'. This is considered an important contributor to catastrophic landslides. Residents' awareness of the importance of nature and beautiful greenery is more pronounced due to the promotional efforts by the Government National Landscape Department under the Min-

istry of Housing and Local Government, established in 1996. [19] examines more comprehensive survey data to suggest environment attitudes are more stable than their earlier literature implied. Establishment in line with the vision of Prime Minister Datuk Seri Dr Mahathir Mohamad, who wants to landscape the necessary aspects in all development projects and responsibilities to state governments and local authorities deal with the serious aspects such as landslides.

The actual amount of landslide damage of a specific landslide event depends on the vulnerability of the affected socio-economic and ecological systems, i.e., broadly defined, on their potential to be harmed by a hazardous event. This indicates, as [20] has pointed out, that investments in human capital has the highest payoff. Residents can be made aware of landslides problems and take action to address them. Solutions and implementation plans largely crafted by the residents work the best. However, outside technical assistance is often needed to find technically sound solutions. In the case of Penang's Program for Landslide Control the solutions were largely determined by the government. However, the residents understood the issues and believed that they could and should do something about them. Hence, the aim of this research is to identify the level of knowledge and awareness aspects among the residents to prevent further landslide and environmental damages occur in the long term plans and to collect information helpful in enhancing residents' knowledge and awareness about the landslide hazards in Penang Island, Malaysia.

3 METHODOLOGY

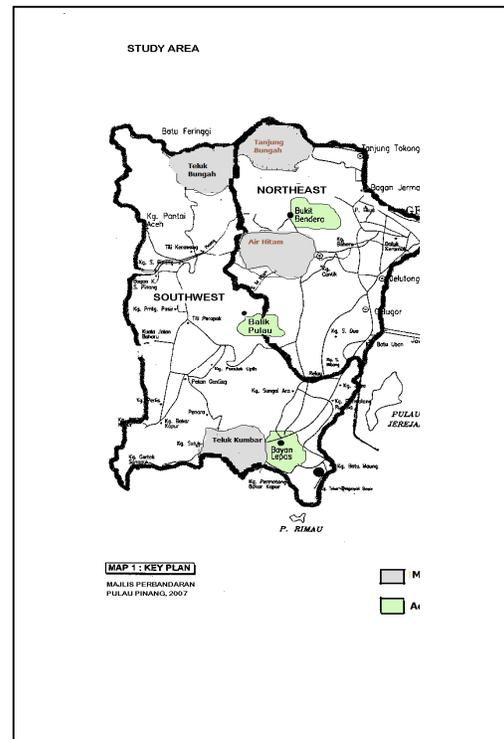
The present study is attempt to examine the landslide awareness of residents in six districts of Penang Island in Malaysia according to residence, gender, age, occupation, race, marital status, monthly salary income and educational background. The research was a guided close-ended questionnaire type of study utilizing survey method. Six districts in Penang Island (one of the 13 states in Malaysia) were randomly selected for the study. Six districts were chosen randomly from northeast (3 districts) and southwest (3 districts). These areas are oftenly exposed to landslides and also exposed to soil erosion. The sample size of the study was between 10 to 15 people randomly selected for each area from the total of 60 people. A return rate of 100% was obtained. Prior to data collection, permission was sought from the residents of their homes. Chi Square Test-test was employed to find out the significance of difference between the level of knowledge and awareness among the residents in the selected areas. SPSS was used for the analysis the data. Landslide awareness refers to the knowledge and awareness of residents holds towards the landslide hazards. It which consisted two dimensions ranging from knowledge and awareness to landslide disasters. There were three parts, firstly about residents' background. The second part contained 15 questions with two choices 'A' and 'B' for each were given. The last part of the questionnaire measured the awareness level of landslide issues, which contained 50 questions. The questions consisted of general awareness of landslide hazards with the response format of a five points Likert scale ranging from strongly agree (5) to distrongly agree (1). Each question is giv-

en a choice of (1) most not agreed (2) not agreed (3) less agreed (4) agreed (5) most agreed to the questions asked. Specifically, 60 residents completed the guided questionnaires. Resident feedback in response to the questionnaire is quite good, but only a few residents that are less proficient in the Malay language, the researcher had to explain about the contents of the questionnaire before they answer the questionnaire easily without any doubt.

4 RESULTS

Chi-Square Test was used to compare the frequencies of knowledge and awareness scores of from different demography background of selected divisions in two districts (southwest and northeast) of Penang Island. There were six areas selected for the sample chi square test, Teluk Bahang, Batu Ferringhi, Bukit Bendera, Bukit Gambir, Paya Terubong and Balik Pulau. Figure 1 shows the study areas of researcher. Researcher study the landslide awareness of the residents which consisting of housewife and the retired. 60 people involved in this study, comprising 10 people from each area. Based on the measurement formula chi square test analysis found a significant relationship between the questions of the level of awareness of surveyed residents about the dangers of landslides. Figure 1 shows the study locations of researcher.

Figure 1 Study Areas



Penang Island Map

65% residents agree to live in the hills for it is comfortable atmosphere compared to 35% them disagree. This indicates

that residents agree to live in the hills are more than the people whom dislike living on hills. This shows that the residents are less aware about the dangers of landslides because they are not concerned with the threat of landslides hitting them at present. 85% residents agree the problem of landslides caused by heavy rains and 93.3% of them agree the main causes of landslides is due to human activities that damage the environment whereas 85% of them do not agree to cut trees along the hillsides. This shows that the residents has a high awareness of landslides as they understand the importance of planting trees along the slopes to prevent landslide. 80% residents blame the government for the landslide occurrences whereas 83.3% of them stated that there is no landslide programs organized by state governments to handle landslides, while 83.3% of them agree that they do not have a place much safer to be protected from the event of a landslide threat. This reflects that residents have a high awareness of the disaster as they realize that the state government has a greater role in managing projects in hills, especially in areas related to social and economic development. Residents also aware that lack of landslide programs and shelter have made them to take serious view of the landslide disaster.

91.7% residents agree that all the residents who live in the hills need to buy insurance for property damage or loss of life in landslides and 93.3% of them agree the state government is fully responsible for the landslide occurrences, while 78.3% residents state that self awareness is very important after facing the threat of landslides. This shows that residents have a high awareness of the disaster as they are more aware of insurance which is a tool compensation for self protection and property. Residents also aware that government itself should play an active role to address landslides and in protecting the people and the environment. 91.7% residents agree to imprison anyone who violates the laws or the hill land policy without binding or withdraw permits construction on hill land. This reflects that residents have a high awareness of landslides and their rights in exploring the hill land. 55% of them agree the government does not require higher spending to restore the landslide occurrences and this reflects the residents have a low awareness of landslides as they are less vulnerable to statistical state government spending in the recovery operation to the usual disaster affected by landslides. 61.7% residents agree with the government's preparation in handling the landslide problem compare to the balance percentage of residents whom are not sure or do not know about the preparedness of government in addressing disaster in this island. This reflects residents are still lack of awareness about the dangers of landslides.

90% residents agree the landslides only occur at high altitudes only. This indicates that the residents have a highly aware of the landslide as they understand the zonal areas of landslides in the state and 68.3% of them do not agree that landslides will occur in the years to come. This implies that the residents are less aware of landslides as they could not determine whether landslides will occur in the future, as many of the hills have been explored and have experienced the effects of the disaster. 83.3% residents were still interested in buying a house on the hill although they knew of the dangers of

landslides. This reflects that residents are still not aware of landslide even though they have much heard and read in the media about the threat of landslides in some areas in Malaysia. 68.3% of total residents agree that landslides occurred slowly, and this shows they are still less aware and less exposed to the processes of the landslide which occurred in the state. 83.3% residents afraid and concerned about the threat of landslides and do not want to live in the areas showing signs of landslides and this shows they are aware of landslide whereas 80% of them agreed that every houses in the high lands of need to install the siren landslide. This shows residents have high awareness of landslides as they are wary of the disasters and demonstrate a precautionary attitude to the future in an effort to distance themselves from the disaster. 93.3% residents agree the construction engineers must be efficient and competent in the selection of housing areas to avoid the threat of landslides in the future and this indicates the residents have a high awareness of disaster as they understand the talent should be skilled in various areas of strategic and able to find solutions to various problems through innovative thinking. 56.7% of them agreed the safety of landslides in Penang is not dangerous. This shows the residents are less aware of this disaster in which is the main center of this disaster after the Selangor.

78.3% residents agree never to feel the impact of landslides and reflects that awareness of the residents are low since they have never experienced such disasters as likely to stay away from the disaster area or less to know about the disaster that has nothing to do with themselves whereas 50% of them had felt the effects of landslides and indicates they have a high awareness of disaster as they had experienced such catastrophic events. 91.7% residents agree every event which felt weird or strange about the landslide will be reported to authorities as soon as possible, and 65% of them agree to plant a variety of trees along the hill slopes, while 93.3% of them agree the government should gazette the hill lands reserves. This reflects the residents have a high awareness of landslide disaster due to stay alert for strange happenings around them and realize the importance of planting the trees along the hill slopes to prevent soil particles falling from high to flat land and gazette the hill areas to explore arbitrarily due to concrete development. 63.3% residents agree and interested in buying a property on the hill after seeing the devastating landslide in the hills and 81.7% of them agree they had never heard of a landslide awareness program in any area in Malaysia, while 68.3% residents agree the government's plan will effectively prevent landslides in the years to come. Residents are still lack of awareness about the disaster as they were not exposed to it and not sure of the effectiveness of government programs in the future.

58.3% of them love to work with anyone in efforts to reduce landslides and 85% of them agree the policy of high land in Malaysia is made to suit the interests of superiors without thinking of the public interest, while 91.7 % of them agree to seek a further more and secure place to live on when aware of the warning signs of landslides in the areas where they live. This indicates the awareness of residents about the landslides is high because they have a higher responsibility, less selfish,

have a high knowledge of the highland policy implementation in Penang. 93.3% residents agree to plant trees at the foot of the hill to avoid the threat of landslides and realize the importance of roots to avoid the clutches of the disaster. This implies the high aware of the residents on landslide whereas 56.7% of them agree not to believe in the predictions of landslide disaster. This reflects that residents have low awareness of the disaster as due to lack trust to the prediction of natural disasters that can help people get ready if there are signs of landslides in their areas. 83.3% residents agree there are strange sounds heard before the occurrence of landslides and this shows the residents have a force high enough stimulus to recognize a sign or effect of landslides that occurred in the vicinity of his residence. 76.7% residents agree the rapid technological advancements make people more aware, efficient and intelligent in carrying out responsibilities for environment and this indicates their low awareness of the disaster as more days landslide occurred in tandem with the increase in human and modern technology that almost destroyed the beautiful environment and harmony, particularly in the hills to meet the appetite or desire of humans with no limits.

90% of the residents agree rapid advances in technology make people careless and negligent in discharging their responsibilities towards the environment. This reflects the awareness of people about the disaster is high as the residents were beginning to feel the cause of the landslide in the world which caused by human greediness and indifference to explore the land for various purposes till the lost of thousands of innocent lives and destruction of property valued at and the government had to issue millions of dollars for the restoration of the normal environmental conditions. Although we have reached millennium era, but landslides continue to happen till man cannot find a point solution to prevent the occurrences of landslides due to greed of man himself in the process of exploring the beautiful environment created by God.

5 DISCUSSION

Analysis of the resident awareness about landslide disaster issues indicated that, they rated their awareness at the 'have some awareness' level with an average score of 63% out of 100%, which is a reasonable result for residents. Some awareness gaps were identified however, particularly in the resident awareness responsibility for environmental damage, stern action should be imposed for that invade the hills, to preserve the hills for future generation for which the average resident scores were in the level of 'have some awareness'. This finding underscores that residents have insufficient awareness about the importance of plants on the hill and along the slopes. Residents are moderately knowledgeable, however, in some landslide issues including hill cutting for development and greed for profit. In fact, landslide problems are among the most urgent problems in Penang for which all researchers are trying to find long-term solutions. The analysis of the relationship between resident background and awareness levels did suggest a significant difference. Resident with high education realized that planting trees on hills are important to prevent landslide compare to those suggest to build many buildings

on the hills. Considering the average rating scores of residents for landslide issues, they were rather low in comparison to their knowledge level of landslide issues which corresponded to the level 'the state government should be firm in encapsulates high land policies for the welfare and safety of people around'. These result, indicate that not enough of landslide knowledge among the residents and that setting up links between the environmental and landslide disaster awareness in public practices have failed and indeed, these give us clues regarding the directions for future research in this context. There is low awareness of residents of landslides where they do not support the issue of 'landslides can disrupt the daily activities of the people', but their awareness level was the lowest for the issue of 'engineers have an efficient and competent construction in the selection of the construction areas to avoid the threat of landslides in future'. This showed that residents have low knowledge on landslide disaster.

Penang has no record of systematic landslide occurrences for reference in future. So, it is recommended that a accurate and systematic records of landslides should be provided by certain parties for future researchers. There was no statistical difference between gender and the rating scores of residents with respect to their awareness level about landslide issues. As for resident responses to the multiple-choice, more than 68% of the residents correctly answered the question about the definition of landslide, and their average rating score about their knowledge of the definition of landslide was 65%, which corresponded to 'highland areas shall be maintained to prevent encroachment by unscrupulous people'. This result does not represent an inconsistency because in a multiple-choice format, they could have easily recognized the definition by glancing at the choices. The resident response to the second multiple-choice question about the main components of landslide revealed their insufficient knowledge of these issues, as only slightly more than 53% of the residents identified the correct answer. The responses of the residents regarding their attendance to any landslide activity and previous programs on landslide were quite encouraging in the sense that, only 20% residents out of 60 stated that they had experienced any occurrence regarding landslide and only 40% residents indicated that they do not experienced landslide. This result strongly supports the hypothesis that landslide awareness is not yet extended among residents in this state.

The analysis of the statements section, which was designed to investigate resident opinions on the importance of the need for and knowledge of landslide, suggested that residents had positive attitudes towards landslide, its importance and its place in programs even though they had insufficient educational background and knowledge levels in landslide. Results also indicated that, the vast majority of residents believed that the landslide concept was important for the present and future, it was of interest to all of residents and more emphasis should be placed on landslide at all levels awareness. They also thought that there was something they could have done to promote landslide awareness programs. This finding suggests that residents were ready to take action for landslide even residents willing to act in addressing the problem of landslides in this country by working with all levels of society.

Then, this study revealed that the residents surveyed had significant awareness gaps in landslide issues and were facing urgent problems regarding their educational background in landslide, despite having a reasonable background in landslide issues in general. Finally, the study has shown that almost all residents in Penang believe that the research awareness of landslides is very important to prevent the recurrence.

ACKNOWLEDGMENT

The author(s) would like to thank the Municipalities of Penang island and the private companies for their moral supports and the cooperation of all residents who are directly involved in the study.

REFERENCES

- [1] Bandura, A. *Self-efficacy*. New York: W. H. Freeman. 1997.
- [2] Gibson, E. and Pick, A.D. *An Ecological Approach to Perceptual Learning and Development*. Oxford: Oxford University Press. 2000.
- [3] Gibson, J.J. *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton-Mifflin. 1966.
- [4] Gibson, J. J. *The Ecological Approach to Visual Perception*. Boston: Houghton- Mifflin. 1979.
- [5] Chan, N.W. "Environmental hazards associated with hill land development in Penang Island, Malaysia: Some recommendations on effective management". *Emerald Disaster Prevention and Management* 7(4): 305-318. 1998.
- [6] Abdullah C.H., Mohamad A., Yusof M.A.M., Gue S.S. & Mahmud M. "Development of Slope Management in Malaysia". 2007.
- [7] Stern, P. C. Psychology, sustainability, and the science of human-environment interactions. *American Psychologist*, 55, 523-530. 2000.
- [8] Silverstein, M. *Disasters*. McClean, VA: Brassey's. 1992.
- [9] Oskamp, S., *Attitudes and opinions*. Englewood Cliffs, NJ: Prentice-Hall, 1977.
- [10] Gue, S. S. & Tan Y. C. "Landslides: Abuses of The Prescriptive Method", *International Conference on slope 2006, Kuala Lumpur, 7 - 8 August 2006*. 2006.
- [11] Bem, D.J. *Beliefs, attitudes and human affairs*. Belmont, California: Brooks/Cole, 1970.
- [12] Mirovitskaya, N. and Ascher, W. *Guide to sustainable development and environmental policy*. London: Duke University Press. 2001.
- [13] Dunlap, R.E. & Dillman, D.A. Decline in public support for environmental protection: Evidence from a 1970-1974 panel study. *Rural Sociology*. 41, 382-390. 1976.
- [14] Keating M. *Agenda for Change : A Plain Language Version of Agenda 21 and Other Rio agreements*, Centre for our Common Future, Geneva, Switzerland, 70 p. 1993.
- [15] Pettus, A. Environmental education and environmental attitudes. *Journal of Environmental Education*. 8, 48-51. 1976.
- [16] Cohen, M.R. Environmental information versus environmental attitudes. *Journal of Environmental Education*, 5,5-8. 1973.
- [17] Young, R.A. The relationship between information levels and environmental approval: The wilderness issue. *Journal of Environmental Education*. 11(3), 25-30. 1980.
- [18] Burris-Bammel, L.L. Information's effect on attitude: A longitudinal study. *Journal of Environmental Education*, 9 (4). 1978.
- [19] Mitchell, R.C. Silent spring/solid majorities. *Public Opinion*. Aug/Sept.,16-20. 1979.

- [20] Thompson, R.L. Incentives for sustainable agriculture and rural development. In (S.A. Brett, ed.) *Integration of Sustainable Agriculture and Rural Development Issues in Agricultural Policy*, pp. 87-99. Winrock International, Morrilton, Arkansas. 1996.

Improved Decision tree algorithm for data streams with Concept-drift adaptation

K.Ruth Ramya, R.S.S.Vishnu Priya, P.Panini Sai, N.Chandrasekhar

Abstract— Decision tree construction is a well studied problem in data mining. Recently, there has been much interest in mining streaming data. Algorithms like VFDT and CVFDT exist for the construction of a decision tree but, as the new examples are added, a new model has to be generated. In this paper, we have given an algorithm for construction of a decision tree that uses discriminant analysis, to choose the cut point for splitting tests thereby optimizing the time complexity to $O(n)$ from $O(n \log n)$. Also various adaptive learning strategies like contextual, dynamic ensemble, forgetting and detector approaches have been analyzed and handling of concept-drift occurred due to gradual change in data-set is discussed using naive Bayes classifier at each inner node.

Index Terms— Adaptive learning strategies, Bayes Classifier, Concept-drift, Data Streams, Decision trees, Discriminant analysis, VFDT.

1 INTRODUCTION

Data streams are problems where the training examples used to construct decision models come over time, usually one at a time. In usual real-world applications the data flows continuously at high-speed. Recently, a new model of data processing focused by the database community is the data streams, in which data arrives in the form of continuous streams [3, 4, 5, 7, 8]. The key issue in mining on streaming data is that only one pass is allowed over the entire data. Moreover, there is a real-time constraint, i.e. the processing time is limited by the rate of arrival of instances in the data stream, and the memory available to store any summary information may be bounded.

In complex systems and for large time periods, we should expect changes in the distribution of the examples. Natural approaches for these incremental tasks are adaptive learning strategies, that is, incremental learning algorithms that take into account concept drift. In this paper we present a framework of an algorithm that generates a tree for data streams and various adaptive learning strategies to detect concept-drift. The main contributions are based on the reduction in the time complexity of proposed algorithm by the use of discriminant analysis - to choose the cut point for splitting function, the use of bayes classifier at each node - to detect concept drift and thereby controlling the online error rate.

2 RELATED WORK

In the literature of machine learning, several methods have been presented to deal with time changing concepts. The related work has been analyzed from two dimensions. One is the concept-drift and the other, improving the splitting criterion function of the tree algorithm.

A concept-drift occurs due to various types of changes seen in distribution of data. Identification of patterns is possible in systematic drift whereas, no patterns can be identified in the case of seemingly unsystematic drift, for a given the available data. Thus an adaptive strategy needs current, labeled data.

For the data that comes from a process having several recurring states, systematic drift refers to a situation in which the current state can be inferred by the available data. Gradual drift refers to slow, continuous change over time. In contrast, sudden or abrupt drift means that substantial drift occurs without prior notice. Sudden drift might well be systematic.

Depending on the type of drift, various adaptive learning strategies can be applied. For example, Incremental learning strategies can be applied, if drift is gradual or when current, labeled data is available. However when there is no availability of current, labeled training data, adaptive strategies are still possible. Such strategies, require the assumption of a specific underlying drift model.

Domingos and Hulten have addressed the problem of decision tree construction on streaming data [8]. Their algorithm guarantees a probabilistic bound on the accuracy of the decision tree that is constructed. In this paper, we revisit the problem of decision tree construction on streaming data using discriminant analysis. Thus, data mining algorithm developed for streaming data also serve as a useful basis for creating approximate, but scalable, implementations for very large datasets.

- K.Ruth Ramya is currently working as Associate Professor in Dept. of CSE, KLCE(Autonomous), Vaddeswaram(AP), India, E-mail: RUTHMOSES.MATHI@gmail.com
- R.S.S.Vishnu Priya is currently pursuing bachelors degree program in computer science and engineering in KLCE(Autonomous), Vaddeswaram(AP), India, E-mail: vishnurayala@gmail.com
- P.Panini Sai is currently pursuing bachelors degree program in computer science and engineering in KLCE(Autonomous), Vaddeswaram(AP), India, E-mail: panini.rckr@gmail.com
- N.Chandrasekhar is currently pursuing bachelors degree program in computer science and engineering in KLCE(Autonomous), Vaddeswaram(AP), India, E-mail: chandrasekhar284@gmail.com

3 PROBLEM DEFINITION

In this section, we focus on the problem of decision tree construction on streaming data. We give a framework of the algorithm below, which will be served as the basis for next sections.

```

ALGORITHM: TREE CONSTRUCTION

Stream Tree (Stream D)
global Tree root, Queue PQ, AQ;
local Node node, Tuple t;
PQ←NULL; AQ←NULL;
add (root, AQ);
while not (empty (PQ) and empty (AQ))
    t←D.get ();
    node←classify (root, t);
    if node ∈ AQ
        add (node.sample, t);
        if node.stop_condition_satisfied
            remove (node,AQ);
        if node.enough_samples()
            use split_function to get the best split;
            (node1, node2)←node.create();
            remove (node, AQ);
            add ((node1, node2), PQ);
        while required_memory_exists (AQ, PQ)
            get (node, PQ);
            add (node, AQ);
    
```

This algorithm uses two queues, PQ and AQ. AQ stands for active queue and denotes the set of currently working decision tree nodes on expanding. PQ is the set of decision tree nodes that have not yet been split, but are not being processed currently. This distinction is made because additional memory is required for actively processing each node. Based on the availability of free memory, the set AQ is constructed from the set PQ by including as many nodes as possible. The algorithm is initiated by putting the root of the decision tree in the set.

The input to the algorithm is a stream of data instances, denoted by D. A data instance t from this stream is successively obtained and the current decision tree node (denoted by node) to which this data instance belongs to is determined. If node belongs to the set AQ, then t is added to the set of samples available to process node. We then check if node satisfies the stop condition. If so, node is removed from AQ. Otherwise, we check if we now have sufficient information to split node. If so, the node is split, removed from AQ, and its two child nodes are added to the set. When both PQ and AQ are empty the algorithm terminates.

This algorithm is exposed to number of issues in decision tree construction on streaming data. One of the issue is, computationally how efficiently we examine all possible splitting conditions associated with a node. The following section gives the details of how this issue is handled.

4 USING DISCRIMINANT ANALYSIS FOR SPLITTING CRITERIA

The algorithm starts with the construction of a single leaf. When a splitting test is installed at a leaf, the leaf becomes a decision node, and two descendant leaves are generated. The splitting test has two possible outcomes: i) True, associated with one branch and ii) False, associated with the other branch. The splitting tests are generated over a numerical attribute and are of the form $attribute_i < value_j$. For all numerical attributes, the most promising $value_j$ is chosen. We use a modified version of the analytical method presented in [4] for split point selection. The only sufficient statistics required here, are the mean and variance per class of each numerical attribute. This is the major advantage over other exhaustive methods used in C4.5 [12] and in VFDTc [2], because all the necessary statistics are computed on the fly. It is the desirable property for huge data streams because it guarantees constant time processing of each example.

The analytical method uses a modified form of quadratic discriminant analysis to include different variances on the two classes and the distribution of the values of an attribute follows a normal distribution for both classes. Let ,

$$\varphi(\bar{x}, \sigma) = 1/(\sigma\sqrt{2\pi}) \exp(-(x - \bar{x})^2/(2\sigma^2))$$

be the normal density function, where \bar{x} and σ^2 are the sample mean and variance of the class. From the sample set of examples at the node, the class mean and variance for the normal density function are estimated. The quadratic discriminant splits the X-axis into three intervals $(-\infty, r_1), (r_1, r_2), (r_2, \infty)$, where r_1 and r_2 are the possible roots of the equation, $p(-)\varphi\{\bar{x}_-, \sigma_-\} = p(+)\varphi\{\bar{x}_+, \sigma_+\}$, where $p(i)$ denotes the expected probability that an example belongs to class i . As we prefer for a binary split, we choose the root closer to the sample means of both classes. Let r be that root. The splitting test candidate for each numeric attribute 'i' will be of the form $Att_i \leq r_i$. To choose the best splitting test from the candidate list we use a heuristic method. We use the information gain to choose, from all the splitting point candidates, the best splitting test. The splitting test with the maximum information gain is chosen. This method only requires the mean and standard deviation for each class per attribute. Both quantities are easily and incrementally maintained.

For the generated tree, to detect concept drift, at each inner node maintain a naive-Bayes classifier [1] trained with the examples that traverse the node. The next section will illustrate the process of detecting the concept drift.

5 CONCEPT-DRIFT

In the real world concepts are often not stable but change with time. The underlying data distribution may change as well. The model built on old data will be necessarily updated. This problem is known as concept drift. The change in the distribution or the concept-drift may be - Sudden, Gradual or Recurring.

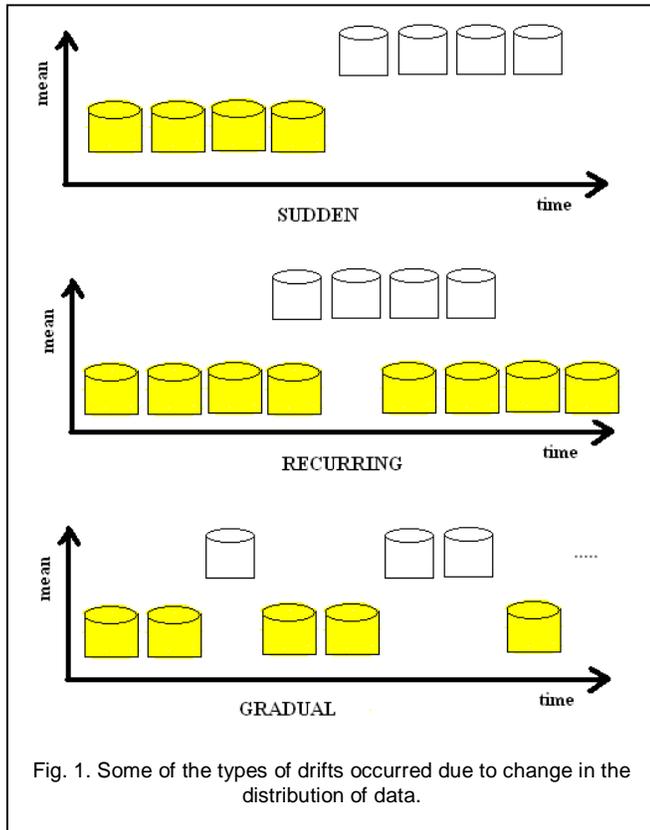


Fig. 1. Some of the types of drifts occurred due to change in the distribution of data.

A. Adaptive learning strategies

There are various strategies for handling the concept-drift. Some of them include Detector approach - detect the change and cut the old data, Forgetting approach - forget old data and retrain at a fixed rate, Contextual approach - build many models, switch models according to the observed incoming data and Dynamic ensemble approach - build many models and dynamically combine them. The following table lists the types of drifts handled by various approaches.

TABLE I
ADAPTIVE LEARNING STRATEGIES FOR HANDLING VARIOUS TYPES OF DRIFTS

SNo.	Adaptive Learning Approach	Handles
1	Detector	Sudden Drift
2	Forgetting	Sudden Drift
3	Dynamic ensemble	Gradual Drift
4	Contextual	Recurring Drift

B. Handling Concept-Drift due to gradual change in data distribution

The basic idea of the drift detection method is to control the online error-rate where in the sequence of examples of the data stream there is a change from one context to another. If the distribution of the examples that traverse a node is stationary, the error rate of naive-Bayes decreases. If there is a change on the distribution of the examples the naive-Bayes error will increase [5]. When the system detects an increase of

the naive-Bayes error in a given node, an indication of a change in the distribution of the examples, this suggests that the splitting-test that has been installed at this node is no longer appropriate and the entire sub tree rooted at that node is pruned.

When a new training example becomes available, it will cross the corresponding binary decision trees from the root node till a leaf. At each node, the naive Bayes installed at that node classify the example. The Binomial distribution gives the general form of the probability for the random variable that represents the number of errors in a sample of n examples. We use the following estimator for the true error of the classification function $p_i = (\text{error}_i / i)$, where i is the number of examples and error_i is the number of examples misclassified, both measured in the actual context. The estimate of error has a variance. The standard deviation for a Binomial is given by $s_i = \sqrt{(p_i * (1 - p_i)) / i}$, where i is the number of examples observed within the present context. For sufficient large values of the example size, the Binomial distribution is closely approximated by a Normal distribution with the same mean and variance. Considering that the probability distribution is unchanged when the context is static, then the $(1-\lambda)/2$ confidence interval for p with $n > 30$ examples is approximately $p \pm \alpha * s_i$. The parameter λ depends on the confidence level.

The drift detection method manages two registers during the training of the learning algorithm, p_{min} and s_{min} . Every time a new example i is processed those values are updated when $p_i + s_i$ is lower or equal than $p_{min} + s_{min}$. We use a warning level to define the optimal size of the context window. The context window will contain the old examples that are on the new context and a minimal number of examples on the old context. Suppose that in the sequence of examples that traverse a node, there is an example i with correspondent p_i and s_i . The warning level is reached if $p_i + s_i \geq p_{min} + 1.5 * s_{min}$. The drift level is reached if $p_i + s_i \geq p_{min} + 3 * s_{min}$.

With this method of learning and forgetting we ensure a way to continuously keep a model better adapted to the present context. The method uses the information already available to the learning algorithm and does not require additional computational resources.

6 CONCLUSION

This paper presents an incremental learning algorithm appropriate for processing high-speed numerical data streams with the capability to adapt to concept drift. The analytical method used is restricted to two-class problems. In future it can be extended to solve problems with more than 2 classes. In [2] the authors presented an extension to VFDT, that which deals with continuous attributes uses a Btree to store continuous attribute-values has complexity $O(n \log(n))$. The complexity of the discriminant analysis method used here is $O(n)$ and thus the efficiency of this algorithm is improved. Also, a brief summary of various adaptive learning strategies were discussed.

REFERENCES

- [1] R.O. Duda, P.E. Hart, and D. Stork. Pattern Classification. New York, Willey and Sons,2001.
- [2] J. Gama, R. Rocha, and P. Medas. Accurate decision trees for mining high-speed data streams. In P.Domingos and C. Faloutsos, editors, Procs. of the 9th ACM SigKDD Int. Conference in Knowledge Discovery and Data Mining. ACM Press, 2003.
- [3] J. Gehrke, F. Korn, and D. Srivastava. On computing correlated aggregates over continual data streams. In Proc. Of the 2001 ACM SIGMOD Intl. Conf. on Management of Data, pages 13–24. acmpress, June 2001.
- [4] Phillip B. Gibbons and S. Tirthapura. Estimating simple functions on the union of data streams. In Proc. of the 2001 ACM Symp. on Parallel Algorithms and Architectures, pages 281–291. ACM Press, August 2001.
- [5] Tom Mitchell. Machine Learning. McGraw Hill, 1997
- [6] S. Guha, N. Koudas, and K. Shim. Data-streams and histograms. In Proceedings of ACM Symp. on Theory of Computing (STOC), pages 471–475. ACM Press, 2001.
- [7] S. Viglas and J. Naughton. Rate-based query optimization for streaming information sources. In Proc. of the 2002 ACM SIGMOD Intl. Conf. on Management of Data, June 2002.
- [8] P. Domingos and G. Hulten. Mining high-speed data streams. In Proceedings of the ACM Conference on Knowledge and Data Discovery (SIGKDD), 2000.
- [9] Dynamic Integration of Classifiers for Handling ConceptDrift, Tsymbal,A., Pechenizkiy, M.,Cunningham,P.&Puuronen,S.Information Fusion, Special Issue onApplications of Ensemble Methods,9(1),pp.56F68,2008.
- [10] Reference Framework for Handling Concept Drift: An Application Perspective. Žliobaitė,I. and Pechenizkiy, M. Technical report, Eindhoven University of Technology,2010
- [11] G. Castillo, "Adaptive learning algorithms for bayesian network classifiers," Ph.D. dissertation, Universidade de Aveiro, Departamento de Matematica, 2006.
- [12] A. Tsymbal, "The problem of concept drift: definitions and related work," Department of Computer Science, Trinity College Dublin, Tech. Rep., 2004.

Solar Energy Utilizing 3-Wheeler Electric Car in Bangladesh

Md. ZakariaMahbub^{1a}, Husnain-Al-Bustam^{2b}, T.M. Iftakhar Uddin^{3c},Md. AbrarSaad^{4c}

¹ Department of Mechanical and Chemical Engineering

^a Islamic University of Technology

e-mail: zakariajon1313@gmail.com

² Department of Electrical and Electronic Engineering

^b Islamic University of Technology

e-mail: husnailb@gmail.com

³ Department of Mechanical and Chemical Engineering

^c Islamic University of Technology

e-mail: tanvir1711@hotmail.com

⁴ Department of Mechanical and Chemical Engineering

^d Islamic University of Technology

e-mail: saad_mce@yahoo.com

Abstract: Bangladesh is one of world most densely populated country with population growth rate 1.566%. For this increasing number of population we need more transport facility. Recently our government has imported a very environmental friendly electric car to this purpose. Its fuel is electricity but it lacks efficiency and speed. At the same time it requires a huge amount of electricity to charge up its battery by which this vehicle could be run nearly ten hours in a day. But we are now in a situation of acute power crisis. And currently the power generation could not meet the actual demand. This newly emerged vehicle is giving rise to the power crisis in a large scale as the number of this kind of car is increasing day by day all over Bangladesh. This paper represents the R and D work to reduce the consumption of electricity from national grid and increase the efficiency of this electric car. To do this we have mounted a 100W solar panel on the roof which is completely vertical. From the result we found that if we use this kind of system with the electric car total electricity consumption from the national grid will be reduced.

Keyword: Electric Car, Electricity, Power Crisis, Roof, Solar Panel, Bangladesh, 100W

1. Introduction

Electricity is a form of energy but not a source. It must be reproduced from some source. All electricity in Bangladesh is reproduced within the country from natural gas, hydro power and a very small amount is generated by oil-using diesel. Reliable & adequate supply of electricity is a pre-condition for poverty reduction and economic development [1]. At present the total power generation capacity of our country is 6727 MW (up to June 15, 2011) [2]. Out of the capacity 3534 MW from the public sector & 3193 from the private sector, this is 53% and 47% respectively of the total generation capacity [1]. A brief view of present generation capacity is given in the Table-1. Actual electricity demand could not be met for the last few years due to the shortage of available generation capacity. But this fact is

also true that the generation capacity has increased slightly. So maximum generation of 2087 MW in 1995-1996, 3218 MW in 2001- 2002, 3458 MW in 2002-2003, 3622 MW in 2003-2004, 3751 MW in 2004-05, 3812 in 2005-06, 3718 in 2006-07, 4130 MW in 2007-08, 4162 MW in 2008-09 and 4,606 MW in 2009-10 could not remove power crisis in the country[1]. Inthe FY 2010-11 maximum generation was 4,890 MW (June 13, 2011). So there is demand for electricity outstrips supply by around 2,000 megawatts at peak hours[1]. So this is the current scenario of the power sector of Bangladesh. This 3-wheeler vehicle is giving rise to the power crisis. The good aspect of this vehicle is that it is ecofriendly and earned a huge popularity especially in the rural areas of Bangladesh. But to charge its battery it requires a huge amount of electricity. The number of this kind of vehicle is increasing day by day as it is cheaper, ecofriendly and could carry four or five passenger at a time. For the increasing numbers of vehicle to charge we need a massive amount of electricity which are not part of total country's power demand. So we can say this demand is extra. To solve this problem we have done research on this issue. First we put a 100W solar panel vertically on the roof of the vehicle. Use of solar panel in country like Bangladesh has introduced a new horizon in power saving. More than half of the year hot sunny days exist. The average temperature here is 26.1⁰C(79⁰F). The specifications of the electric car are given below in the table-2. And we have done a calculation to determine the amount of saving electricity which is about 4.9 ampere out of 17 to 20 ampere required to run the motor. So the use of solar panel on the vehicle's roof is highly appreciable which is environmentally friendly and economic. Our main theme of this research work is to reduce the consumption of electricity which in turn will reduce the power crisis of our country.

TABLE-1
PRESENT GENERATION CAPACITY

S/N	Public Sector	Generation Capacity (MW)
1	BPDB	2620
2	APSCL	659
3	EGCB	255
	Subtotal	3534 (53%)
S/N	Private Sector	Generation Capacity (MW)
1	IPPs	1271
2	SIPPs (BPDB)	99
3	SIPPs (REB)	226
4	15 YR. Rental	168
5	3/5 YR. Rental	441
	Quick Rental	988
	Subtotal	3193 (47%)
	Total	6727

TABLE-2
Specification of the Electric Car

Battery	5 Batteries (12 Volt X 5= 60 Volt)
Motor	1000W
Electricity Consumption	3-4 Units, 220 Volt
Break System	3-4 Mechanical
Mileage	(80-100) km

2. SYSTEM ARCHITECTURE

The main component of our system is the 100W solar panel. The average weight of this panel is approximately 10 kg[3]. We mounted this panel on the roof of the electric car as depicted in the figure-1. The specification of the solar panel is given below[3].



Figure 1. Experimental Setup

TABLE-3

SPECIFICATIONS[10]

100W Solar Panel Specifications (Test condition: 1000W/m ² , AM1.5, 25°C)	
Max Power(Watt)	100w
Tolerance	+3%
Size(mm)	1200*532*40
Weight(kg)	10
Number of Cells	36* Mono crystalline cells
Size of Cells(mm)	125*125
Max Power Voltage(V)	17.2
Max Power Currencty(A)	5.80
Open Circuit Voltage(V)	21.6
Short Circuit Currencty(A)	6.60
Max System Voltage(V)	1000
Wind Bearing	60m/s (200kg/sq.m)
NOCT (Nominal operating cell temperature)	47°C(+2°C)

This solar panel was purchased online and it is CE certified. The block diagram of our system is depicted in the figure 2.

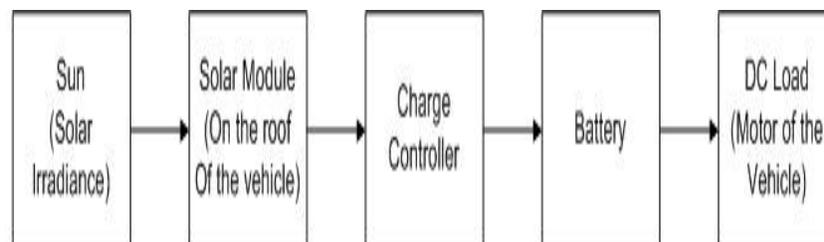


Figure 2. Block Diagram

3. RESULT ANALYSIS

3.1. GRAPH

3.1.1. TEMPERATURE AND CURRENT VS TIME GRAPH

Readings are taken in the month if March, April, May and June. We have taken the readings of three days from 8 am to 5 pm. Temperature and the current output from the solar panel at different day time is shown in the graph. The electricity output varies with the change of temperature at different times. It is shown that the average current is about 4ampere from the solar panel.

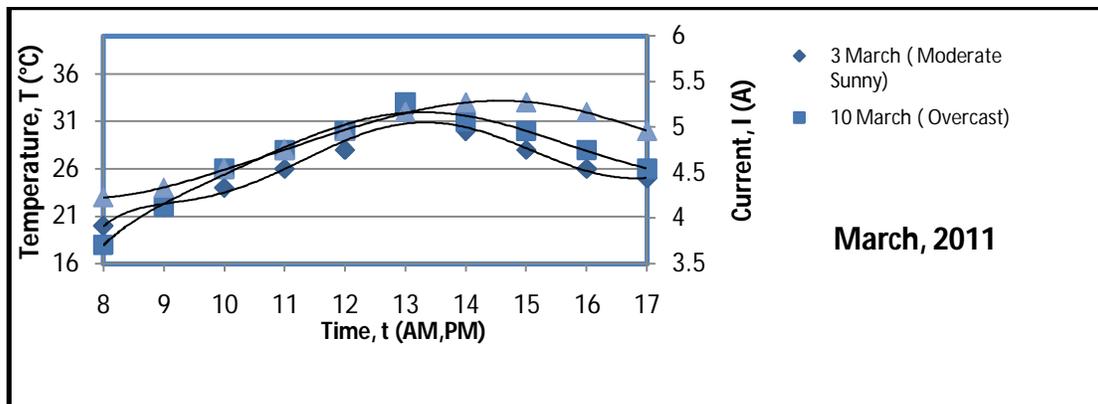


Figure 3. Temp(°C) & Current(Amp) vs. Time Curve for March 2011.

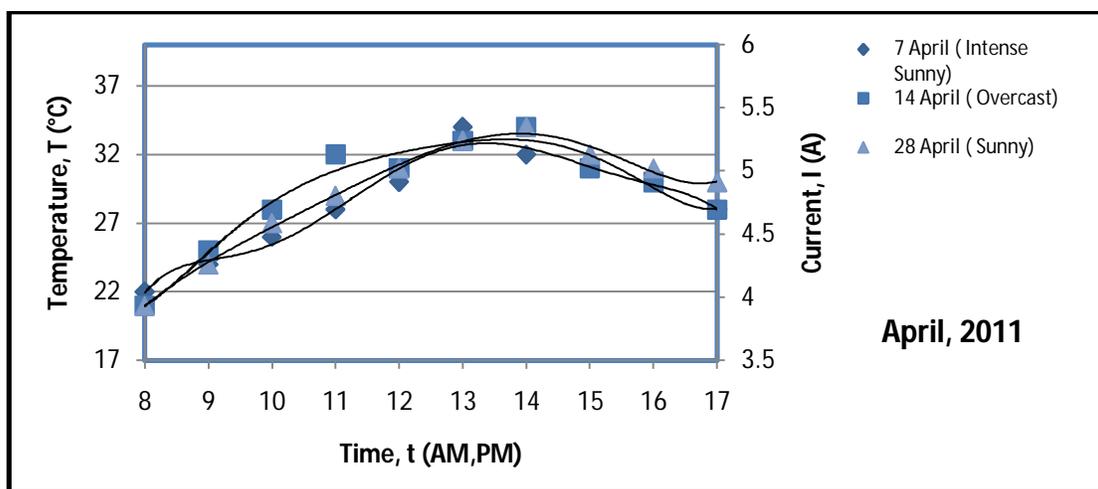


Figure 4. Temp(°C) & Current (Amp) vs. Time Curve for April 2011.

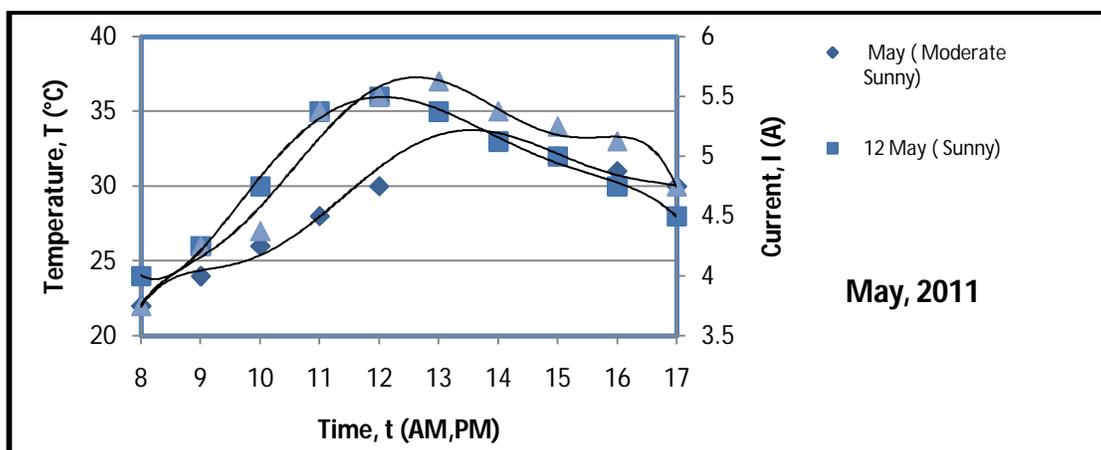


Figure 5. Temp(°C) & Current (Amp) vs. Time Curve for May 2011.

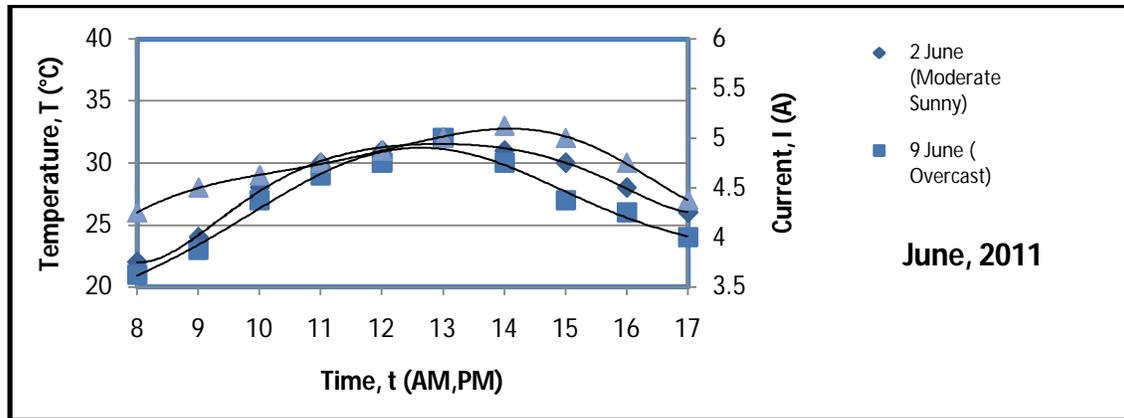


Figure 6. Temp(°C) & Current (Amp) vs. Time Curve for June 2011.

3.1.2. POWER SUPPLIED VS PARTICULAR DAYS GRAPH

For this the readings are taken for four months (March, April, May and June). The graph between particular days vs power supplied shown below. It actually signifies the condition with and without solar panel. From the experiment we have seen that if we use 1000W solar panel at best it is capable of producing 790W. That means from a single electric car we can save this amount of power. For example if there are 15,000 electric cars in all over Bangladesh, and if we use this system we will be able to save $(15,000 \times 790) = 11850000W$ or 11.85 MW electricity.

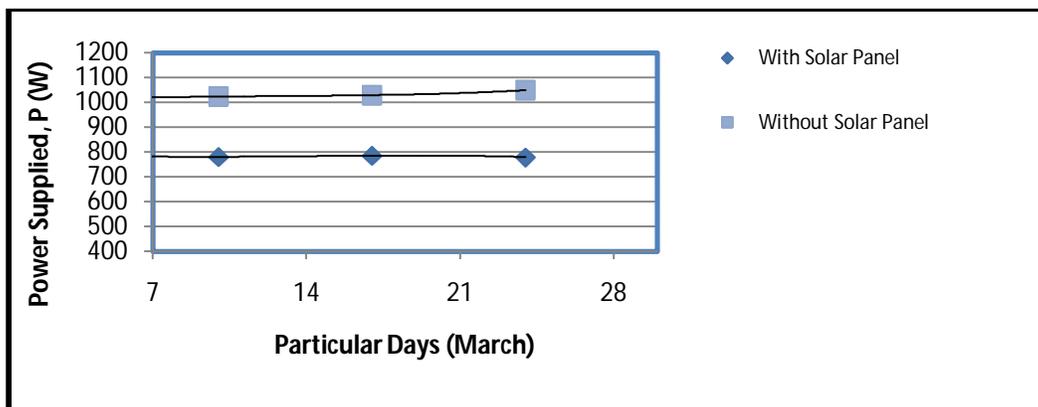


Figure 7. Power supplied (W) vs. Particular Days Curve for March 2011.

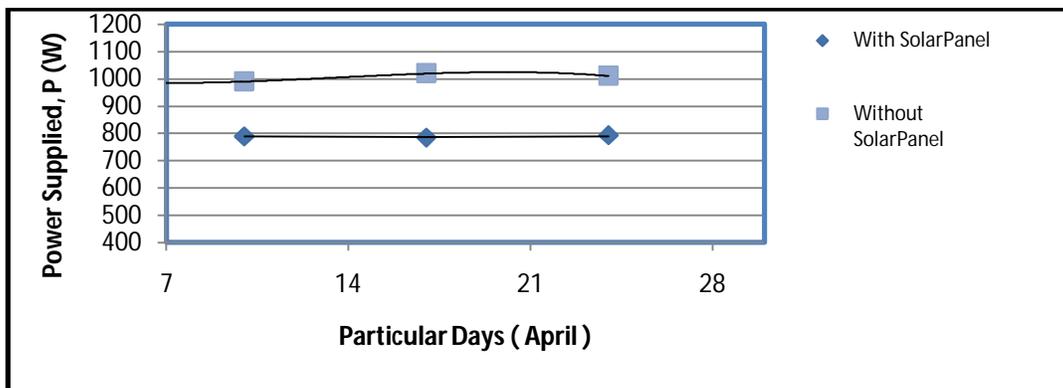


Figure 8. Power supplied (W) vs. Particular Days Curve for April 2011.

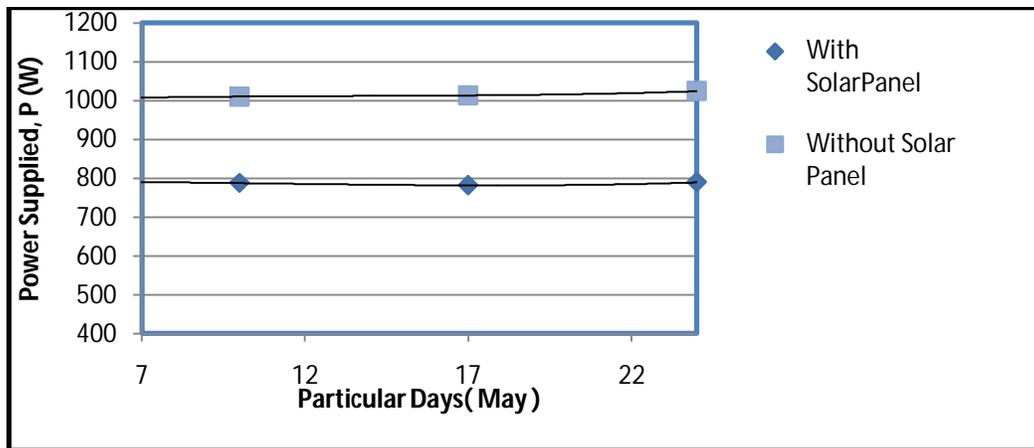


Figure 9. Power supplied (W) vs. Particular Days Curve for May 2011.

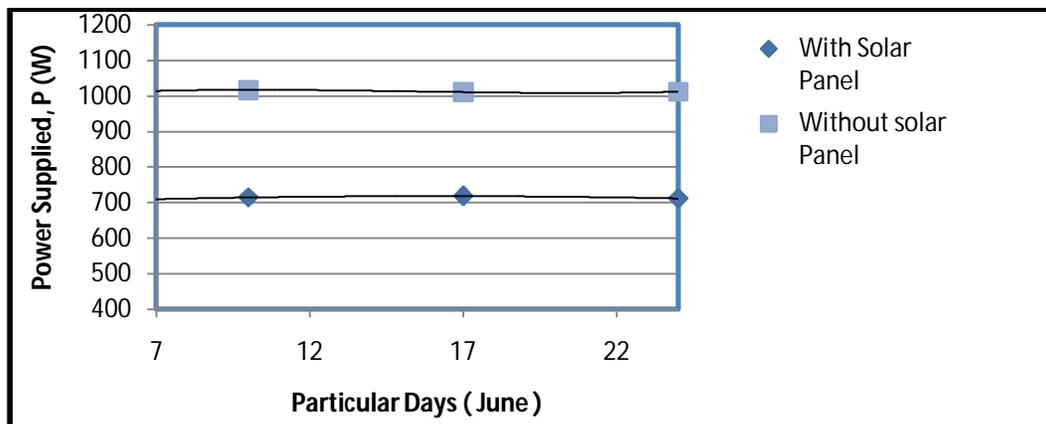


Figure 10. Power supplied (W) vs. Particular Days Curve for June 2011.

4. CALCULATION

4.1. PANEL SIZING

Panel Specification:

Power: 100W

V(Open Circuit) : 32.5 V

V (MPPT): 27.14 V

4.2. CALCULATION

1. 15KWH x 6 hr = 90KWh
2. 90 KWH / 4.5hr = 20 KW
3. 20 KW / 200 W = 100 Panel
4. no of panel in series in 1 set: $360 \text{ v} / 27.12 \text{ v} = 11.79 = 12$ panel in series
5. $200 \text{ w} \times 12 \text{ pieces} = 2.4 \text{ KW}$
6. no of set in parallel = $20 \text{ KW} / 2.4\text{KW} = 8.33 = 8$ panel in parallel

4.3. BATTERY CALCULATION

Battery voltage: 12 Vdc

Battery capacity: 100 Ah

1. Battery sizing: $288 \text{ Vdc} / 12 \text{ Vdc} = 24$ pieces in 1 set [Here inverter input voltage is 288v]
 2. 1 set = 24 pieces x 12V x 100 Ah = 28.8 KWh
 3. $90 \text{ KWh} / .8 = 112.5 \text{ KWh}$
 4. $112.5 \text{ KWh} / 28.8 \text{ KWh} = 3.90 = 5$ set
- Total Battery: 5 x 24 pieces= 96 pieces
So 24 pieces Battery are in series and 5 set are in parallel

4.4. ENERGY SAVED

Current required to run the motor of easy bike = 18A.

Current obtained from the solar panel = 4A (avg).

So energy or current saved = $(4/18) \times 100 = 22.22\%$

7. CONCLUSION

From the experimental result analysis we have found that using of solar panel on the roof of the electric car could be possible solution of the power crisis in Bangladesh. Actually government import these electric car's engine only. The chasis are made with the help of local technology. So it is possible to build such kind of vehicle with solar panel. Besides the price of the fuels on the rise and the vehicle run by these fuels are not environmentally friendly. On the otherhand electric car are totally environmentally friendly. So we should adopt this system.

REFERENCE

1. KhairulAnam, Husnain-Al-Bustam, "Power Crisis & Its Solution through Renewable Energy in Bangladesh", Journal of Selected Areas in Renewable and Sustainable Energy (JRSE), September Edition, 2011, pp, 13-19.
2. ZahariTaha, Rossi Passarella, JamaliMdSah&Nasrudin Bin AbdRahim (2008) "A Review on Energy Management System of a SolarCar", The 9th Asia Pacific Industrial Engineering & Management Systems Conference,3-5 December 2008, Nusa Dua, Bali, paper 93,pp. 2527-2530.
3. Shimizu Y., Komatsu Y., Torii M., Takamuro M. (1998), Solar CarCruising Strategy and Its Supporting System, JSAE Review 19, 143-149.
4. Z. Taha, R. Passarella, J.M. Sah, H.X. Hui, N. Ahmad, R.A.R. Ghazilla, Y.H. Jen, and T.T. Khai,"Developing of TelemetryMonitoring system for a Solar Vehicle", Proceeding of the

IASTED International Conference on Solar Energy, 16-18 March 2009, phuket, 647-035.pp
44-50.

5. Bangladesh Diesel Plant (BDP), Gazipur. Dhaka, Bangladesh.
6. R. Edinger and S. Kual, Renewable Resources for Electric Power: Prospects and Challenges. Quorum Books, London, 2000.
7. Renewable energy report was prepared by Prof. Dr. Neem Chandra Bhowmik, Director Renewable Energy Research Centre, University of Dhaka under a consultancy assignment given by the Asian and Pacific Centre for Transfer of Technology (APCTT).
8. Renewable Energy (2004) Volume: 29, Issue: 5, Publisher: Elsevier, Pages: 643-660, ISSN: 09601481, DOI: 10.1016/j.renene.2003.10.002
9. http://www.banglapedia.org/httpdocs/HT/R_0175.HTM
10. <http://www.brightechsolar.com/solar-panel-100-watt.html>

Analysis of Photovoltaic River to Storage Pumping System by Using the PVSYST Software

Husnain-Al-Bustam, Md. Zakaria Mahbub, M.M. Suvro Shahriar, Md. Imran Hossain, MD. Wasequul Islam, Mirja Shafiqur Rahman

Abstract— This paper represents the analysis of a standalone photovoltaic river to storage pumping system for the rural areas of Bangladesh. Bangladesh is riverine country. SO irrigation for the cultivation is not a great deal of problem here. From the case study it has been demonstrated that most of the cultivable lands are beside any river, lake or canal. The main object of this research work is to develop such a system that could suck water from the river and store into the tank. And that water would be used for the irrigation. One significant reason of this research work is that most of the farmers are irrigate their land by using the river water manually which is cumbersome and time consuming job and not efficient. We have developed a mathematical model of our proposed system and used PVSYST software for the modeling of the “River to Storage Pumping System”. Though this is a completely computational model of the system we believe that we would be able to build this system within \$100.

Index Terms—Analysis, Bangladesh, Mathematical, River, Pumping, Storage Tank, PVSYST, Software, Computational Modeling

I. INTRODUCTION

Irrigation affects various facets of socio-economic issues of country. Agricultural irrigated land refers to agricultural areas purposely provided with water, including land irrigated by controlled flooding [2]. Asian agricultural economies have experienced significant upswings in productivity since the Green Revolution in the 1960s. As widely acknowledged, the Green Revolution was made possible through coordinated infusions of vital inputs into the agricultural systems of the region. Irrigation through major and medium canal systems was one of the most crucial inputs to the Green Revolution [1]. Three decades down the line, the Asian economies benefiting from the Revolution have made famines history. However, after the euphoria over this significant achievement subsided, a major gap was identified—irrigation projects have failed to make the desired difference in the lives of its poor, subsistence and marginal farmers [1]. So we think our project our proposed system will help the farmers to get rid of this

problem of irrigation. Bangladesh is considered as a developing economy which has recorded GDP growth above 5% during the last few years. Microcredit has been a major driver of economic development in Bangladesh and although three fifths of Bangladeshis are employed in the agriculture sector, three quarters of exports revenues come from garment industry. Bangladesh is endowed with plenty of surface water and groundwater resources. Surface water inflows of the country vary from a maximum of about 140,000 cumec in August to a minimum of about 7,000 cumec in February. The alluvial aquifer systems of Bangladesh are some of the most productive groundwater reservoirs. Groundwater in Bangladesh occurs at a very shallow depth where the recent river-borne sediments form prolific aquifers in the floodplains. In the higher terraces, the Barind and Madhupur tracts, the Pleistocene Dupi Tila sands act as aquifers. In the hilly areas, the Pliocene Tipam sands serve as aquifers. The groundwater table over most of Bangladesh lies very close to the surface and fluctuates with the annual recharge discharge conditions.

Total rivers including tributaries and distributaries are about 700 under three mighty river systems: GANGES-PADMA RIVER SYSTEM, BRAHMAPUTRA-JAMUNA RIVER SYSTEM and SURMA-MEGHNA RIVER SYSTEM. Rivers of the southeastern hilly region are considered as the CHITTAGONG REGION RIVER SYSTEM. Principal rivers are: GANGES, PADMA, BRAHMAPUTRA, JAMUNA, SURMA, KUSHIYARA, MEGHNA, KARNAFULI, OLD BRAHMAPUTRA, ARIAL KHAN, BURIGANGA, SHITALAKSHYA, TISTA, ATRAI, GORAI, MADHUMATI, KOBADAK, RUPSA-PASHUR, FENI. So from the above discussion it is clear that Bangladesh is blessed with a lot of rivers. Now the main challenge is that we have made the best utilization of this natural resource for the irrigation purpose. This is why we have we are now in the research of developing cost effective “River to Storage Pumping System”. The system will utilize the photovoltaic module to drive to motor for the suction of the water from the river and store it into the storage tank. From there the water will be distributed in the irrigation land. We have proposed here the storage pump as the farmers would be able to use the water in the night time or in the gloomy day for the irrigation purpose. This paper actually represents the computational modeling of the “River to Storage Pumping System”. First we have developed the mathematical model and

then we have simulated the whole system by using the PVSYST software. The diagrams and tables that are represents in the paper are generated while doing the simulation in the software. In the last portion of the paper we have shown the rivers map for Bangladesh.

II. MATHEMATICAL MODELING

The following mathematical equations are used for the simulation of the total system.

$$I_T = S_H * ((1 - K_D) * R_D + 0.5 * (1 + \cos \beta) * K + 0.5 * (1 - \cos \beta) * \rho) \dots (1)$$

$$T_C = (T_A + (T_{NOCT} - T_S) * (I_O / I_L)) \dots (2)$$

$$\eta_m = \eta_p * \eta_{ec} * \eta_{im} * \eta_{NOCT} \dots (3)$$

$$\eta_{ec} = \eta_c * \eta_{MIS} \dots (4)$$

$$\eta_T \dots (4)$$

$$\eta_c = \eta_o + (1 - \alpha * T_C - T_A) \dots (5)$$

$$\text{Area (m}^2\text{)} = \text{Load} / (\eta * I_D) \dots (6)$$

$$Q \text{ (Kwh)} = (C + \text{NSR}) * \text{Load} / (\text{DOD} * \eta_B) \dots (7)$$

$$\text{DOD} = 1 - \text{SOC} \dots (8)$$

$$I_T - I_D = M * S \dots (9)$$

$$X_{PV} = (X_0 * ((1 + g_o) / (k - g_o)) * (1 - ((1 + g_o) / (1 + k_o))^N)) \dots (10)$$

$$R_{PV} = b * Q * (1 - s) * \sum_{i=1}^{NR} ((1 + g_1) / (1 + k_1))^i \dots (11)$$

Here,

- B= the tilt angle of the array
- ρ = the ground reflectivity
- S_H = the total solar irradiation in a horizontal surface
- K_D =the diffuse insolation factor
- R_D = the daily direct radiation factor
- T_{NOCT} = the nominal operating cell temperature
- T_A =the ambient temperature
- T_S = the standard temperature
- I_O = the standard irradiation
- I_L = the local irradiation
- η_{im} =the illumination mismatch efficiency
- η_{NOCT} = the nominal operating cell temperature

- η_p = the module packing density
- η_{ec} = the encapsulated cell efficiency
- η_{MIS} =the electrical mismatch
- η_T = the optical transmission efficiency
- η_c = the cell efficiency
- η_o = the bare cell efficiency at standard conditions
- α = the temperature coefficient of the cell
- T_C =the operating temperature of photovoltaic modules
- η = the system efficiency
- I_D = the average monthly insolation
- DOD= the depth of discharge
- η_B = the battery efficiency
- C= the number of consecutive days
- M= the balancing parameter
- g_o = the escalation rate for operations & maintenance (O&M)
- K= the money interest rate
- N= the system life in years
- X_0 =the O&M cost in the first year
- b= the battery unit cost
- Q= the nominal battery capacity
- S= per unit salvage value of replaced battery

III. COMPUTATIONAL MODELING

A. Well Characteristics

Level Depth	8 m
Pump Depth	15 m

B. Storage Tank

Volume	20 m ³
Diameter	3.0 m
Water full height	2.0 m
Feeding attitude	5.0 m

C. Hydraulic Circuit

Pipe choice	PE20(3/4")
Piping length	18m
Number of elbow	0
Other friction losses	0.05

D. Water Needs

Water needs avg.	5 m ³ /day
Yearly water needs	1785 m ³
Yearly head avg.	38 meterW
Hydraulic energy	188979 W
PV needs(very roughly)	638 226 W

E. Pump(s) Model and Layout

Sort pump by power 98W, 6-70 m
Nominal voltage = 24V

Nominal current = 7A

F. PV Array

Sort model by power = 60Wp, 24V
 Array nominal power = 240Wp
 Array voltage(50°C) = 29.7V
 Array = 7.1 A

TABLE-1

PV Module : W*H	0.425*0.249 m ²
Module in Series	2
Total Number of Strings	2
Total Number of Modules	4
Total Area	2.1 m ²
Total Power	0.2 kWp

G. PV Array Summary

4 PV module of 60Wp
 Array oper. (50°C,1000W/m2)
 Pump=211W, Vmpp=30V

H. Operating System

Fixed DC-DC converter

I. Pump Characteristics

Pump: 1 pump
Type: deep well pump
Motor type: DC motor
Nominal operating condition:
 Pressure =32.0 meterW
 Flow rate=1.6m3/h

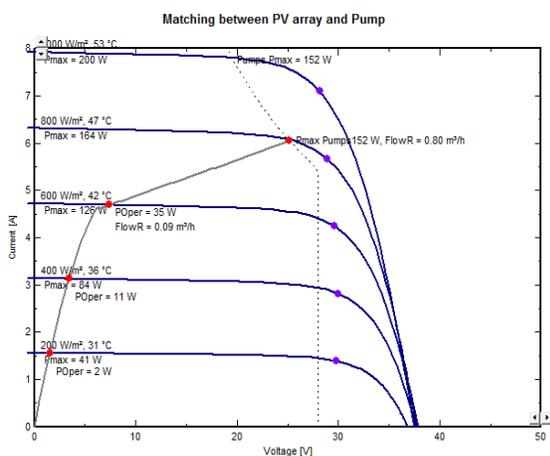


Figure.1 Matching Between PV Array and Pump

IV. SYSTEM DESIGN

The Climate is Sub-tropical monsoon in Bangladesh. Average maximum and minimum winter temperatures are 29°C and 11°C respectively; average maximum and minimum summer temperatures are 34°C and 21°C respectively. Annual

RAINFALL 1,194 mm to 3,454 mm. Highest humidity 80% to 100% (August-September), lowest 36% (February-March)[2]. The following figures (2-6) represent various geographical conditions of Bangladesh and hence we can say we are now in an ideal position for utilizing the Solar Energy.

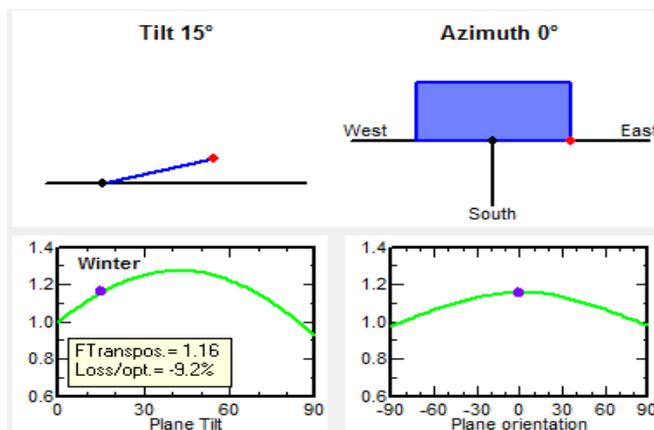


Figure.2. Sun Orientation (Fixed Titled Plane)

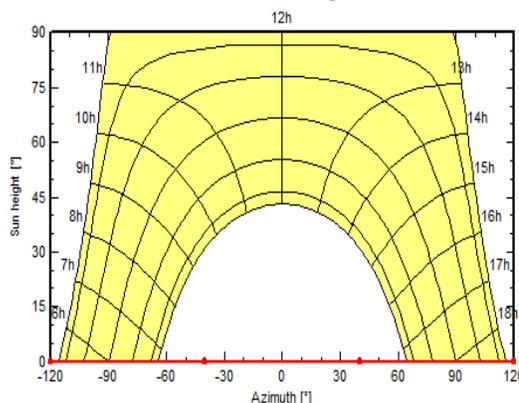


Figure.3. Horizon Line in Bangladesh

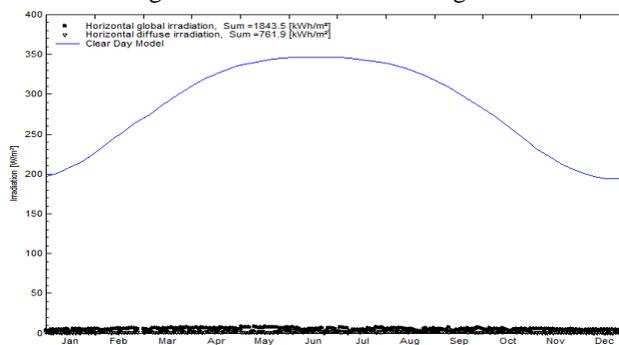


Figure.4. Meteo for Bangladesh Synthetic Data

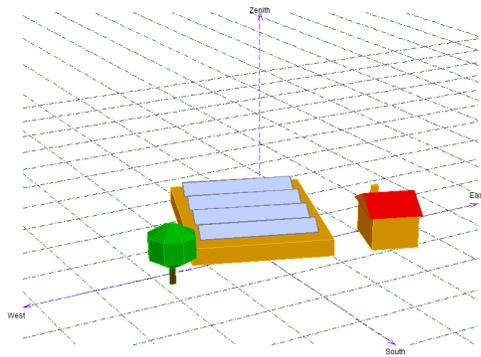


Figure.5. Near Shading of Bangladesh (View-1)

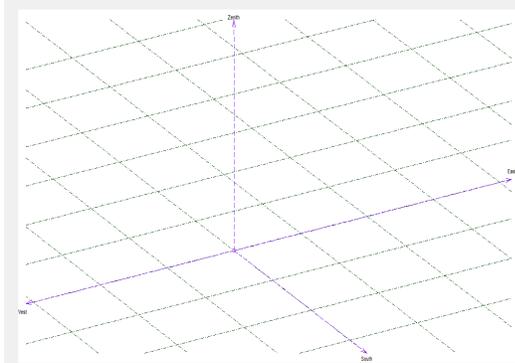


Figure.6. Near Shading of Bangladesh (View-2)

Figure-7 and figure-8 represents the total system architecture. Actually these figures are generated according to the parameters that are defined in the computational modeling section. We actually use the system of array configuration. Figure-9 represents the performance graph of simulated system.

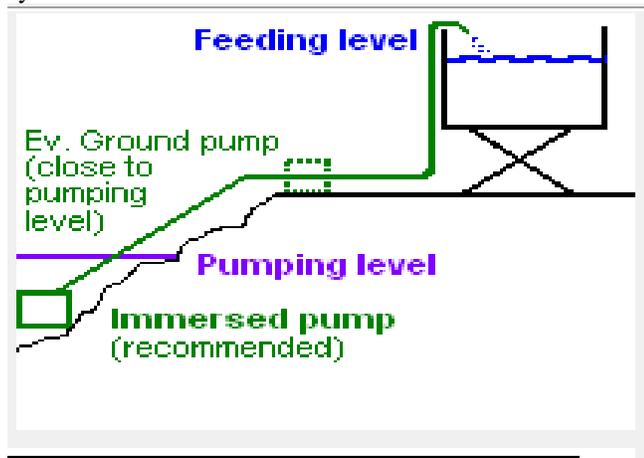


Figure.7. System Architecture

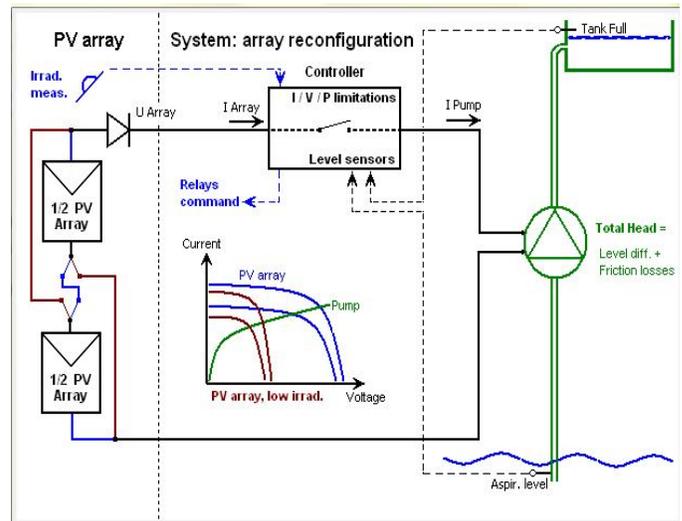


Figure.8. Circuit Diagram of the System

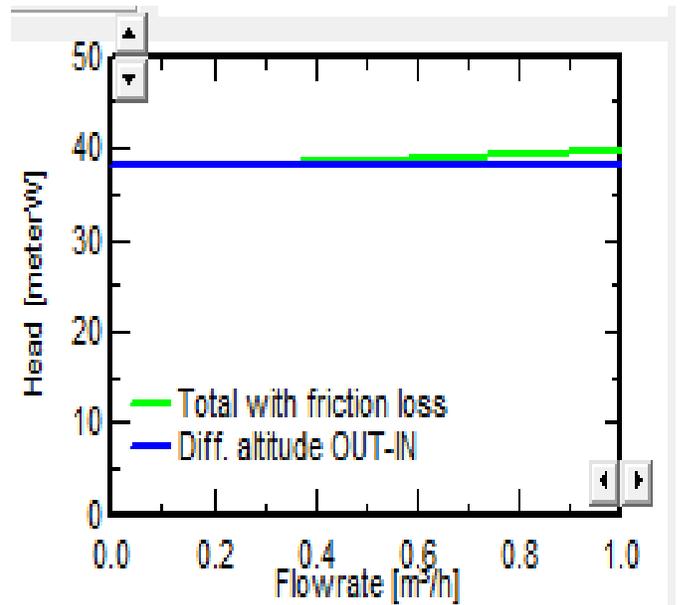


Figure.9. System Performance Overview

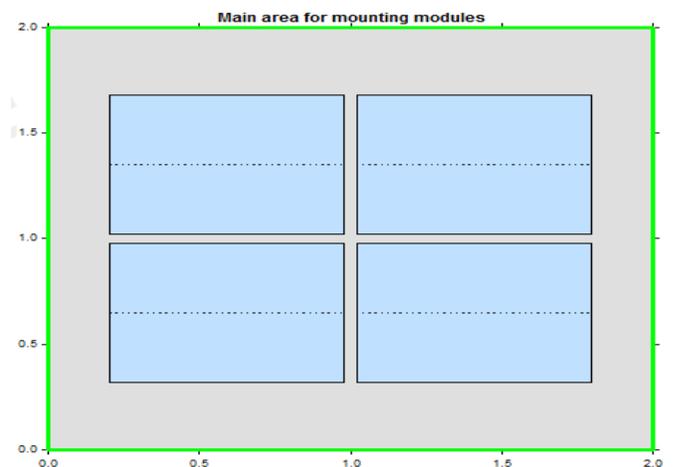


Figure.10. Main Area of Mounting Module

V. PERFORMANCE ANALYSIS

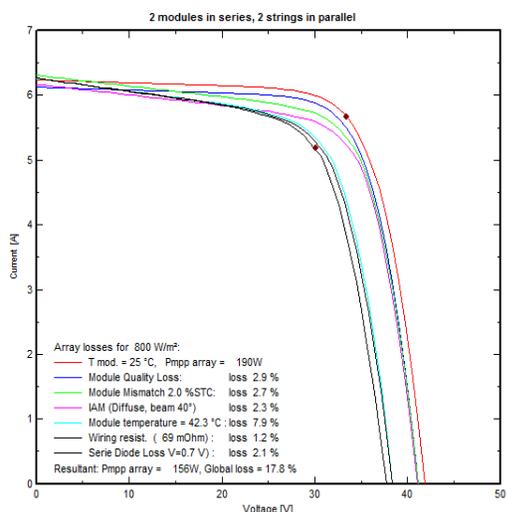


Figure.11. PV Array Behavior for Each Loss Effect

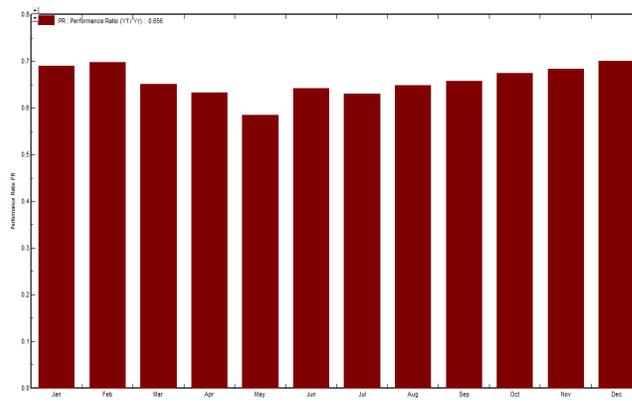


Figure.14. Performance Ratio

VI. TABLES

TABLE-2
BALANCES AND MAIN RESULT

5 m³/day, H=32+6 m, PV=240Wp, Pumps: 2 x 98W, direct coupling

Balances and main results

	GlobEff kWh/m ²	EArMPP kWh	E PmpOp kWh	ETKFull kWh	H Pump meterW	WPumped m ³ /day	W Used m ³ /day	W Miss m ³ /day
January	149.6	29.61	25.55	0.000	39.13	4.428	4.686	0.314
February	160.0	30.96	27.59	0.000	39.21	5.323	4.848	0.152
March	181.7	34.45	29.25	1.338	39.20	5.087	5.000	0.000
April	176.4	33.39	27.61	1.946	39.12	4.979	5.000	0.000
May	195.2	36.50	28.21	3.756	39.28	4.961	5.000	0.000
June	136.0	25.84	21.69	0.000	38.81	3.873	4.369	0.631
July	142.5	26.92	22.30	0.000	38.88	3.885	3.793	1.207
August	141.2	26.82	22.75	0.000	38.80	3.904	3.971	1.029
September	143.2	27.11	23.35	0.000	38.92	4.208	4.061	0.939
October	150.3	28.73	25.12	0.000	38.94	4.369	4.513	0.487
November	160.4	30.84	27.16	0.000	39.19	4.884	4.883	0.117
December	161.0	31.83	27.98	0.000	39.14	4.851	4.750	0.250
Year	1897.5	363.00	308.55	7.040	39.05	4.557	4.571	0.429

TABLE-3
METEO AND INCIDENT ENERGY

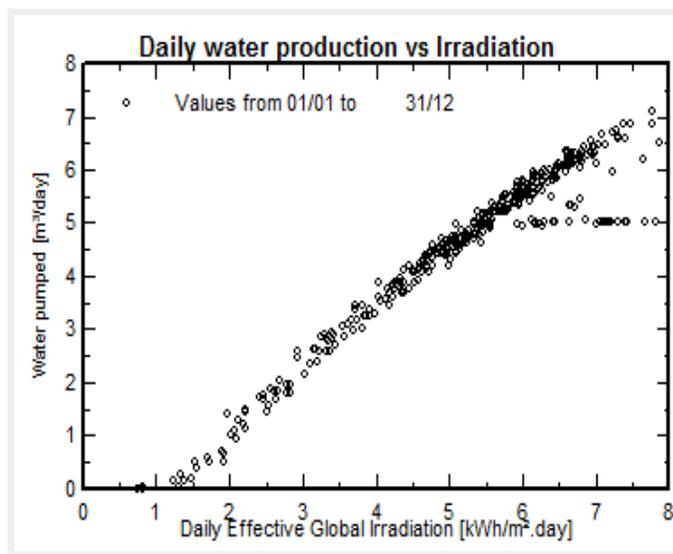


Figure.12. Daily Water Production Vs. Irradiation

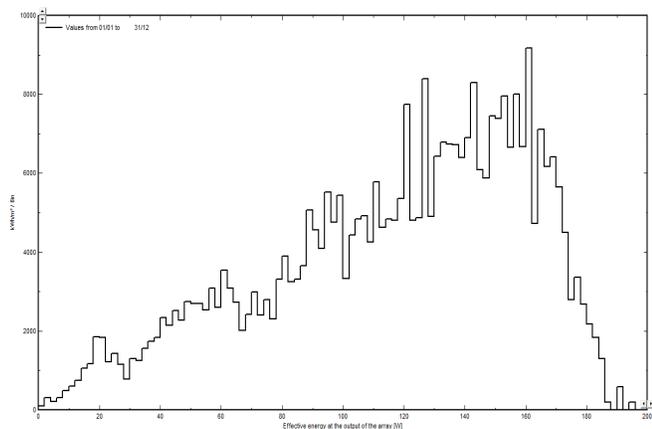


Figure.13. Array Power Distribution

5 m3/day, H=32+6 m, PV=240Wp, Pumps: 2 x 98W, direct coupling

Meteo and incident energy

	GlobHor	DiffHor	T Amb	WindVel	GlobInc	DiffInc	Alb Inc	Diff/GI
	kWh/m ²	kWh/m ²	°C	m/s	kWh/m ²	kWh/m ²	kWh/m ²	
January	128.6	45.61	19.40	1.5	154.4	49.15	0.438	0.318
February	144.1	46.51	21.50	1.7	164.8	49.11	0.491	0.298
March	175.4	60.86	26.20	2.1	187.2	61.94	0.598	0.331
April	180.5	72.77	28.10	2.4	181.8	72.09	0.615	0.396
May	209.6	60.87	28.70	2.3	201.2	58.94	0.714	0.293
June	146.8	81.67	28.30	2.3	140.8	79.51	0.500	0.565
July	153.2	80.82	28.80	2.4	147.4	78.77	0.522	0.534
August	148.0	85.04	28.70	2.1	146.1	83.61	0.504	0.572
September	143.2	75.74	28.60	1.9	148.0	75.61	0.488	0.511
October	141.9	65.51	27.60	1.7	155.3	67.05	0.483	0.432
November	138.7	44.25	23.70	1.6	165.5	47.84	0.473	0.289
December	133.5	39.95	20.40	1.4	166.4	44.38	0.455	0.267
Year	1843.5	759.59	25.85	1.9	1959.1	768.00	6.281	0.392

TABLE-4
LOSS IN THE PV SYSTEM

5 m3/day, H=32+6 m, PV=240Wp, Pumps: 2 x 98W, direct coupling

Losses in the PV system

	ModQual	MisLoss	OhmLoss	EArrMPP	MPPLoss	EArUfix	EArray
	kWh	kWh	kWh	kWh	kWh	kWh	kWh
January	1.016	0.657	0.315	29.61	0.741	28.87	28.87
February	1.061	0.686	0.351	30.96	0.467	30.50	30.50
March	1.182	0.765	0.392	34.45	0.357	34.09	34.09
April	1.149	0.743	0.358	33.39	0.265	33.12	33.12
May	1.254	0.811	0.415	36.50	0.272	36.23	36.23
June	0.896	0.579	0.228	25.84	0.209	25.63	25.63
July	0.933	0.603	0.244	26.92	0.202	26.72	26.72
August	0.929	0.601	0.240	26.82	0.217	26.60	26.60
September	0.937	0.606	0.261	27.11	0.239	26.87	26.87
October	0.992	0.641	0.283	28.73	0.262	28.47	28.47
November	1.059	0.685	0.343	30.84	0.390	30.45	30.45
December	1.091	0.706	0.344	31.83	0.757	31.07	31.07
Year	12.500	8.083	3.774	363.00	4.378	358.62	358.62

TABLE-5
LOSS IN THE PUMPING SYSTEM

5 m3/day, H=32+6 m, PV=240Wp, Pumps: 2 x 98W, direct coupling

Losses in the pumping system

	EArrMPP	MPPLoss	EPmpThr	EPmpOvr	E_PmpAv	ELowLev	ETkFull	E PmpOp
	kWh							
January	29.61	0.741	0.812	0.000	25.55	0.000	0.000	25.55
February	30.96	0.467	0.391	0.000	27.59	0.000	0.000	27.59
March	34.45	0.357	0.570	0.000	30.59	0.000	1.398	29.25
April	33.39	0.265	0.689	0.000	29.56	0.000	1.946	27.61
May	36.50	0.272	1.221	0.000	31.96	0.000	3.756	28.21
June	25.84	0.209	1.126	0.000	21.69	0.000	0.000	21.69
July	26.92	0.202	1.535	0.000	22.30	0.000	0.000	22.30
August	26.82	0.217	1.038	0.000	22.75	0.000	0.000	22.75
September	27.11	0.239	0.795	0.000	23.35	0.000	0.000	23.35
October	28.73	0.262	0.683	0.000	25.12	0.000	0.000	25.12
November	30.84	0.390	0.724	0.000	27.16	0.000	0.000	27.16
December	31.83	0.757	0.558	0.000	27.98	0.000	0.000	27.98
Year	363.00	4.378	10.143	0.000	315.59	0.000	7.040	308.55

VII. CONCLUSION

From the mathematical modeling and computational modeling, studying the agricultural and economic scenario of Bangladesh we can conclude that our proposed system could be a perfect system for the suction of water from the river for irrigation purpose. The storage tank is used so that the farmers would be able to irrigate their land in the night time. And at last we would like to mention that we are now in the fabrication stage of this research project.

REFERENCES

- [1] Arindam Mukherji, Pro-poor Intervention Strategies in Irrigated Agriculture: Lessons from the Asian Experience, International Waste Management Institute, Website: http://www.iwmi.cgiar.org/publications/Water_Policy_Briefs/PDF/WPB28.pdf
- [2] http://www.banglapedia.org/httpdocs/HT/B_0141.HTM
- [3] <http://www.tradingeconomics.com/bangladesh/agricultural-irrigated-land-percent-of-total-agricultural-land-wb-data.html>
- [4] Luis Santos Pereira, Theib Oweis, Abdelaziz Zairi, Irrigation management under water scarcity, [Agricultural Water Management](#), Volume 57, Issue 3, 30 December 2002, Pages 175–206.
- [5] Robert C. Johansson*, Yacov Tsurb, Terry L. Roec, Rachid Doukkalid, Ariel Dinar, Pricing irrigation water: a review of theory and practice, *Water Policy* 4 (2002) 173–199
- [6] Renewable energy report was prepared by Prof. Dr. Neem Chandra Bhowmik, Director Renewable Energy Research Centre, University of Dhaka under a consultancy assignment given by the Asian and Pacific Centre for Transfer of Technology (APCTT).
- [7] *Renewable Energy* (2004) Volume: 29, Issue: 5, Publisher: Elsevier, Pages: 643-660, ISSN: 09601481, DOI: [10.1016/j.renene.2003.10.002](https://doi.org/10.1016/j.renene.2003.10.002)
- [8] Ahmed. I; ET AL.: How much Energy Does It Take To Make a Gallon of Soy Diesel? National Soy Diesel Development Board, 2009
- [9] <http://www.tradingeconomics.com/bangladesh/fossil-fuel-energy-consumption-percent-of-total-wb-data.html>
- [10] <http://www.energybangla.com/index.php?mod=article&cat=Petrroleumsector&article=24944>
- [11] http://en.wikipedia.org/wiki/Geography_of_Bangladesh
- [12] <http://www.pvsyst.com/>
- [13] <http://www.pvsyst.com/en/publications/meteo-data-sources>
- [14] Solanki, Chetan Singh. "Renewable Energy Technologies: Practical Guide for Beginners" 1st Edition, Prentice-Hall of India Pvt.Ltd, ISBN :8120334345

Loss diagram for "5 m³/day, H=32+6 m, PV=240Wp, Pumps: 2 x 98W, direct coupling" - year

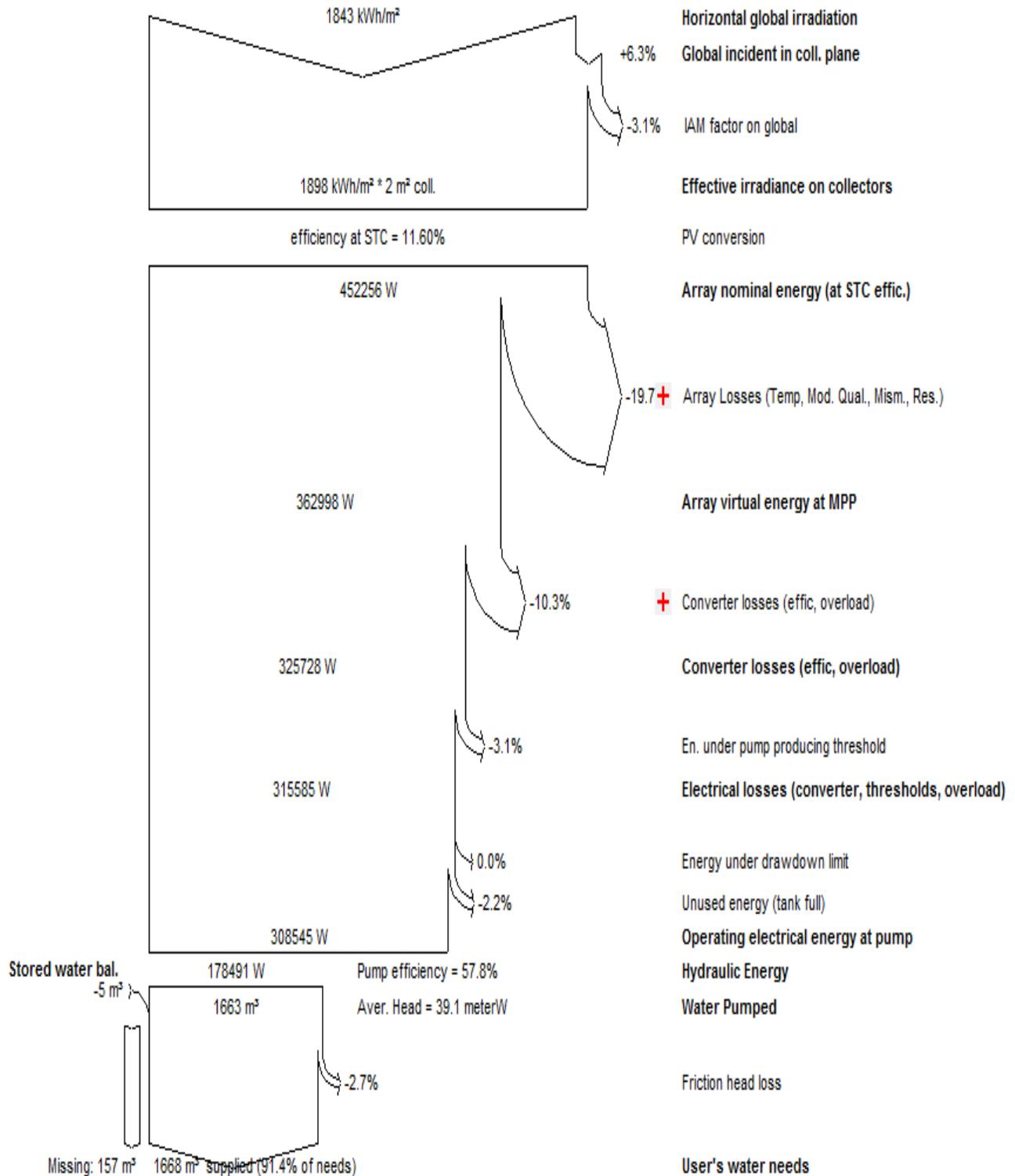


Figure.12. Loss Diagram for the System

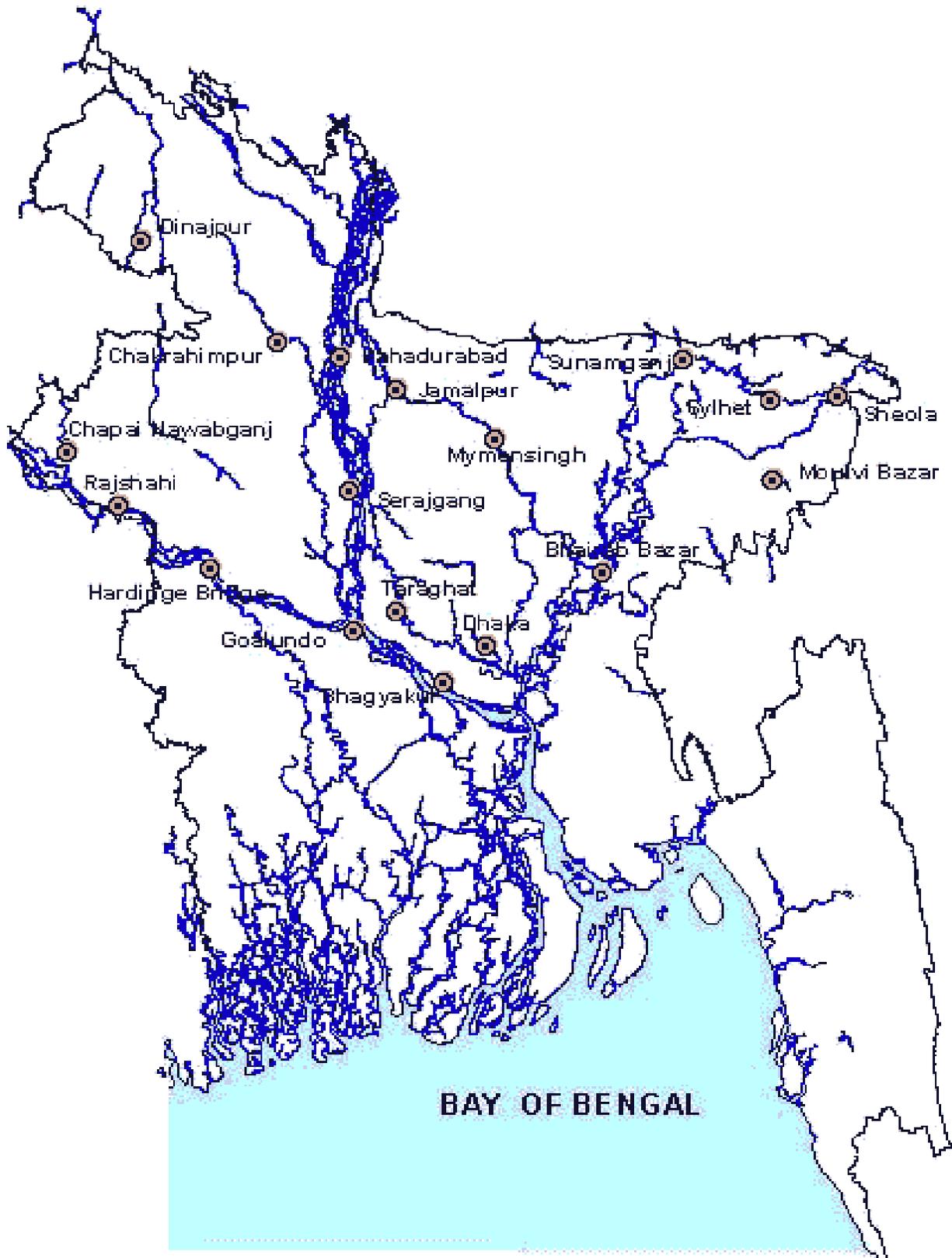


Figure.13. Bangladesh's Rivers Map

Speed of convergence. CEE and Western Balkans countries

Enton Duro

Abstract — The implications of neoclassical model in growth theory are still dominant. One of the implications is convergence hypothesis which means that in the long run income per capita will converge in the steady state level. With regards it is suggesting a natural methodology for finding support of convergence hypothesis. Different researches found the speed of convergence for different “convergence club” countries and regions around the world. Regarded to this it is suggested “convergence club” countries for Albania which is composed by Central, Eastern Europe (CEE) and Western Balkans countries because of same similar characteristics. The observed period for this study is 2000-2010 and the explanatory variable is GNI per capita. In this study, after testing the convergence hypothesis in cross-sectional data set through regression analyses, it is estimated the speed of convergence around 2% per year between this “convergence club” country.

Index Terms— “Convergence club” for Albania, Convergence of income per capita, GNI, Mathematical neoclassical model of growth, Speed of convergence.

1. INTRODUCTION

Albania and other similar countries are in the process of European integration, process of which determines an achievement of certain requirement. There are different countries with similar historical facts, countries. Countries like Estonia, Latvia, Lithuania, Poland, Czech Republic, Slovakia, Hungary and Slovenia in May 2004, and in January 2007 Romania and Bulgaria, are now part of EU. Other countries like Albania, Macedonia, Serbia, etc are trying to make possible this participation. Being part of the EU required obligation of achieving determinant objectives in different spheres of life and one of them are economic criteria. Candidate and potential candidate countries need to converge in common objectives. For this reason it is looking for convergence of income per capita between these countries.

First part of this study illustrated the concept of convergence which will be tested. Literature review about pro and cons convergence hypothesis is suggested in the second part. Then, in the third part, mathematical model of convergence explain how to pose the convergence problem and how to quantify the response of them. The fourth part of the study trying to constructs “convergence club” countries for Albania according similarities between them. The last part represents an empirical finding by cross-sectional regression analyses in SPSS 17.

2. THE CONCEPT OF CONVERGENCE

Beyond the history neoclassical model was mostly considered by the researchers to predict convergence in term of income per capita. The pioneers of neoclassical model are Solow [16], [17] and Swan [20]. Considering diminishing return of physical capital in production, which is the crucial assumption in contest of convergence, they conclude that a closed economy will converge in its steady state in term of income per capita, or more detailed, in the presence of exogenous

technological progress income per capita will grow only by the rate of technological progress. In the neoclassical model steady state level depends on the depreciation rate of physical capital, the growth rate of population and the rate of saving. Differences in these country indicators, which are considered constant in time, saving rate, population growth and depreciation, make the differences in steady state level. On the basis of the neoclassical assumptions endogenous saving rate taking in consideration optimal consumption choice by individual in time and the result is the same, in the long run an economy will converge in steady state level of income per capita [3], [7], [19]. This means that each country, in the long run, will converge at its steady state level. So, lower the level of income per capita, compared to steady state level, higher should be rate of growth for achieving long run equilibrium.

Researches considering these theoretically frameworks have constructed a convergence terminology. *Absolute* convergence refers to the process by which relatively poor countries grow faster than rich ones. The notion of absolute convergence maybe implies that county indicators are the same for all, or more exactly, these indicators are dynamically related to the economic growth and evolve together. Countries may have a common steady state and, in the long run, they can grow at the same rate, rate of exogenous technological change. Reasons for absolute convergence should be: (i) lower level of capital means higher marginal returns which push increasing investments and, in the same time, income to steady state level, (ii) the contagion of poor economies from successful economic model of richer can push them towards steady state level. *Conditional* convergence implies that a given country converges to its steady state and this level depends on the individual county indicators. By this definition it is not necessary to argue that poor counties grow faster than rich ones. How largest is the rate of growth of a given country depends by the distance of its own steady state. This mean that poor countries can have a lower growth rate even that they level of income per capita is lower than rich counties. Reasons for conditional convergence should be: (i) Even though a higher return of capital in poor countries, they cannot increase investment because of higher necessity to consume and, in the open economy, be-

• Enton DURO is currently pursuing PhD program in Economics in University of Tirana. He is assistant of professor in macroeconomics in department of Economics in Faculty of Economy UT. E-mail: duro7enton@gmail.com

cause of higher risk and uncertainty that characterize these countries which is called "poverty trap" in literature, (ii) even though successful model are available poor countries do not have "the capacities" to adopt them. However, conditional convergence permits to analyze, after controlling for steady state differences, a negative relation between growth and initial level of income per capita in the same way of absolute convergence.

3. LITERATURE REVIEW OF CONVERGENCE

The most important assumption that makes possible a presence of convergence in theory is diminishing return of capital. If this condition is not held than income per capita for a given economy do not converge and consequently even the world. In fact, for better explanation of growth emerged new theory of endogenous growth. Romer [11], [12] and Lucas [15] use an excellent theoretical framework to explain that diminishing return of capital is not held in general. Based in AK model, through external effect of knowledge, R&D and innovation, Romer explains that a given country will not converge in term of GDP per capita. Lucas, considering human capital performance and learning by doing in the production process which leads in increasing returns to scale, interprets the same conclusion of non convergence.

The sustention of convergence comes at first by Abramovitz [10] who explains the negative relation between initial level of productivity and growth rate of productivity in the long run. He defines "catch up" the process which make possible that poor can catch rich in term of productivity. Baumol [21] using Maddison data [1], introduced the term of "convergence clubs". The idea of clubs is that if a homogenous sample of countries can be founded, than according to neoclassical model absolute convergence should be observed. He found a negative relation between average rate of growth and initial level GDP per work hour during the period 1870-1979 by the equation $g = 5.25 - 0.75 \ln(Y_{1870})$ $R^2 = 0.88$ for 16 countries called them clubs. Making attention of the ratio of standard deviation from the mean of GDP per work-hour, Baumol concluded that it is fallen quite steadily. Mankiw, Romer and Veil using augmented Solow model by human capital, founded a significant tendency toward convergence for OECD sample [8]. After controlling for rate of saving, population growth and human capital they explained that convergence hypothesis is relevant. One of the regression results offered by them is the speed of convergence around 2% per year. The contribute in convergence hypothesis is offered by Barro and Sala-I-Martin where they have investigated the US data set for 48 states and have suggested a 2% annum as a speed of convergence between them [13]. The same rate of convergence was suggested by Sala-I-Martin from Japanese prefectures [23]. However, absolute convergence it is not present in the world. Mankiw, Romer and Veil conclude that unconditional convergence is not present in their data. Sala-I-Martin did not find absolute convergence between 110 countries for the period 1960-1990 [23]. The set of countries in the world did not converge in the sense of beta convergence. Barro and Sala-I-Martin suggested that

growth rate from 1960 to 2000 is not related with the Log of GDP per capita in 1960 using cross-country data for 112 countries; even more they found a positive relation intending a presence of divergence [14].

This literature review suggests that, with more probability, the hypothesis of conditional convergence is more present in the world. Conditional convergence into the world or absolute convergence between "convergence club" and regions make reasonable the agreement of similarities of them. The diffusion of technology, the influence of social mentality and their behavior, the political condition and decision, the representation of democracy by institutions, openness of economy, and all of that unexplored in growth theories, tend to be more similar in these kind of "convergence club" or regions.

4. MATHEMATICAL NEOCLASSICAL MODEL OF GROWTH WITH CONSTANT SAVING RATE

Assuming Cobb Douglas production function with labor augmenting technological progress [2]. Technological progress transforms the labor input in effective labor input $Y = F(K; AL)$. In the Cobb Douglas function production by labor and capital inputs is given

$$Y = K^\alpha (AL)^{1-\alpha} \quad (1)$$

Where technology and labor input are supposed to be

$$\left. \begin{aligned} A(t) &= e^{gt} \\ L(t) &= L_0 e^{nt} \end{aligned} \right\} \Rightarrow AL = e^{gt} L_0 e^{nt} = L_0 e^{(n+g)t}$$

This means that this factor grows constantly in time by $n+g$. In terms of capital per effective worker we have

$$\tilde{k} = K(t) / L_0 e^{(n+g)t} \Rightarrow K(t) = \tilde{k} L_0 e^{(n+g)t}$$

Where the first derivate per time (t) should be equal to saving in a closed economy

$$\dot{K} = \dot{\tilde{k}} L_0 e^{(n+g)t} + \tilde{k} (n+g) L_0 e^{(n+g)t} = s F[\tilde{k}; L_0 e^{(n+g)t}] \quad (2)$$

Divided per effective worker we have $\dot{\tilde{k}} + \tilde{k}(n+g) = s f(\tilde{k})$. If we take in consideration depreciation rate by (δ) than we take

$$\dot{\tilde{k}} = s f(\tilde{k}) - (n+g+\delta)\tilde{k} \quad (3)$$

In the Cobb Douglas case we have

$$\dot{\tilde{k}} = s \tilde{k}^\alpha - (n+g+\delta)\tilde{k} \quad (4)$$

At steady state capital per effective worker will not change, so $\dot{\tilde{k}} = 0$. If we consider output per effective worker equal to $\tilde{y} = \tilde{k}^\alpha$ in the steady state physical capital per effective worker will be taken by $s \tilde{k}^\alpha = (n+g+\delta)\tilde{k}$. By transformations we get steady state level of capital per effective worker which is

$$\tilde{k}^* = \left(\frac{s}{n+g+\delta} \right)^{\frac{1}{1-\alpha}} \quad (5)$$

And output per effective worker converges in

$$\tilde{y}^* = \left(\left(\frac{s}{n + \delta + g} \right)^{\frac{1}{1-\alpha}} \right)^\alpha \tag{6}$$

The right side of the expression is composed by constants. We have assumed population, technology and depreciation are changed constantly in time. This formulation explains why steady state level should be different for different countries. Output per capita (effective worker) in steady state depends on saving rate, population growth, technological change and depreciation. To consider growth rate of output we can write

$$\tilde{y}^* = \left[\frac{Y(t)}{A(t)L(t)} \right] = \left(\left(\frac{s}{n + \delta} \right)^{\frac{1}{1-\alpha}} \right)^\alpha \Rightarrow Y(t) = \left(\left(\frac{s}{n + \delta} \right)^{\frac{1}{1-\alpha}} \right)^\alpha A(t)L(t)$$

If we like to have growth rate of per capita output we can follow

$$\left[\frac{Y(t)}{A(t)L(t)} \right] = \left(\left(\frac{s}{n + \delta} \right)^{\frac{1}{1-\alpha}} \right)^\alpha \Rightarrow y = \frac{Y(t)}{L(t)} = \left(\left(\frac{s}{n + \delta} \right)^{\frac{1}{1-\alpha}} \right)^\alpha A(t)$$

For facilities we consider

$$\phi = \left(\left(\frac{s}{n + \delta + g} \right)^{\frac{1}{1-\alpha}} \right)^\alpha$$

Growth rate of output will be equal to

$$\dot{Y} = \frac{\dot{A}(t)L(t) + A(t)\dot{L}(t)}{\phi A(t)L(t)} = \frac{\dot{A}L(t) + \dot{L}A(t)}{A(t)L(t)} = \frac{\dot{A}}{A(t)} + \frac{\dot{L}}{L(t)} = g + n \tag{7}$$

And growth rate of per capita output will be

$$\dot{y} = \frac{\dot{\phi}A(t)}{\phi A(t)} = g \tag{8}$$

This formulation demonstrates what we have mentioned before. For a given country growth rate in steady state converges in population growth plus technological change or, growth rate of output per capita converges in technological change. These are the conclusions of neoclassical model.

4.1. SPEED OF CONVERGENCE

Consider equation (4) we can find rate of change of capital per effective worker in any time by

$$\dot{\tilde{k}}/\tilde{k} = s \left(\frac{\tilde{y}}{\tilde{k}} \right)^\alpha - (n + \delta + g) \tag{9}$$

This may be written as

$$\dot{\tilde{k}}/\tilde{k} = s\tilde{k}^{-(1-\alpha)} - (n + \delta + g) = se^{-(1-\alpha)\ln \tilde{k}} - (n + \delta + g) \tag{10}$$

So, the rate of change of capital per effective worker declines as its value converges to steady state. Here we can introduce the concept of speed of convergence. It measure how much the rate of change decreases as its level increases in proportional sense (i.e. by 1%). By formulas it can be written

$$\beta \equiv - \frac{\partial \left(\frac{\dot{\tilde{k}}}{\tilde{k}} \right)}{\partial \ln \tilde{k}} \tag{11}$$

This term is called beta convergence. The negative sign is justified by the fact that does not make sense negative speed of convergence if the capital per effective worker does not achieve steady state. The rapport of the formula is negative because of the opposite movement of the components multiplied per minus makes β positive. We can obtain them by considering absolute value of the first derivate of the right side of 1.10 by $\ln \tilde{k}$

$$\beta = |-(1-\alpha)s\tilde{k}^{-(1-\alpha)}| \tag{12}$$

Speed of convergence is negatively related to the level of capital per effective worker. This means that, when we are at lower level of steady state, speed declines as we approach to the steady state, so β is not constant. In the equation (12) we can substitute the value of capital per effective worker in steady state given by equation (5) and then, after simple transformations, we have

$$\beta^* = (1-\alpha)(n + \delta + g) \tag{13}$$

This is the steady state for β . Speed of convergence, in this equation, represents how rapidly capital per effective worker approaches to the neighborhood of steady state level. So, we can write

$$\frac{\dot{\tilde{k}}}{\tilde{k}} \approx -\beta^* \left[\ln \left(\frac{\tilde{k}}{\tilde{k}^*} \right) \right] \tag{14}$$

Consider Cobb Douglas of (1) we can be written

$$\frac{\dot{\tilde{y}}}{\tilde{y}} = \alpha \left(\frac{\dot{\tilde{k}}}{\tilde{k}} \right) \text{ or } \ln \left(\frac{\dot{\tilde{y}}}{\tilde{y}} \right) = \ln \left(\alpha \left(\frac{\dot{\tilde{k}}}{\tilde{k}} \right) \right)$$

Using (11) and (14) it can be written

$$\frac{\dot{\tilde{y}}}{\tilde{y}} \approx -\beta^* \left[\ln \left(\frac{\tilde{y}}{\tilde{y}^*} \right) \right] \tag{15}$$

Equation (15) says that growth rate of output per effective worker to the neighborhood of steady state, is equal to the speed of convergence multiplied by the gap between steady state and actual level. Again, if countries have the same steady state level poor country will grow faster than rich by the catch up process. The solution of this differential equation can give the equation of testing convergence hypothesis. Equation (15) can be written

$$(\ln \tilde{y})' = -\beta [\ln \tilde{y} - \ln \tilde{y}^*] \tag{16}$$

Substituting $\ln \tilde{y} = z$ we have $\dot{z} = -\beta(z(t) - z^*)$. Here $z_0 = z(0)$ and z^* is steady state. For more we have

$$\dot{z} + \beta z(t) = \beta z^* \tag{17}$$

Considering integrating factor $\mu(t) = e^{\int \beta dt} = e^{\beta t}$ and multiplying both sides

$$e^{\beta t} \dot{z} + e^{\beta t} \beta z(t) = e^{\beta t} \beta z^* \tag{18}$$

Integration of (18) gives

$$z(t) = z^* + Ce^{-\beta t} \tag{19}$$

Initial condition of starting point will have that $z_0=z(0)$ and $z(0) = z^* + C$ where $C = z(0) - z^*$ (PPP).

Differential solution will be
$$z(t) = z^* + [z(0) - z^*]e^{-\beta t} \quad (20)$$

Substitution of $\ln(y)=z(t)$ give
$$\ln \tilde{y} = [1 - e^{-\beta t}] \ln \tilde{y}^* + e^{-\beta t} \ln \tilde{y}_0 \quad (21)$$

Here $\ln \tilde{y}_0$ is the natural logarithm of initial level of output per effective worker. Subtract in both parts of (21) by $\ln \tilde{y}_0$ we have

$$\ln \left[\frac{\tilde{y}_t}{\tilde{y}_0} \right] = [1 - e^{-\beta t}] \ln \frac{\tilde{y}_t}{\tilde{y}_0} + [1 - e^{-\beta t}] \ln \tilde{y}^* \quad (22)$$

In this equation we can see that growth rate is explained by initial level of income per capita and by steady state level of income per capita, all in term of effective worker. If we assume that technology is exogenous and growth by constant rate g , the equation (22) can be formulated in the same way in term of income per capita adding the constant g . Also $[1 - e^{-\beta t}] \ln \tilde{y}^*$ is no longer an explanatory variable if we consider the absolute convergence. Passing from the deterministic definition of hypothesis to the stochastic process we have to consider a normal distribution of the random disturbance in the process by the term u_i . For the given interval of time $[0;T]$ we can write

$$\frac{1}{T} \ln \left[\frac{\tilde{y}_{i;T}}{\tilde{y}_{i;0}} \right] = a - \frac{[1 - e^{-\beta T}]}{T} \ln \tilde{y}_{i;0} + u_{i;0,T} \quad (23)$$

Here $a = g + [1 - e^{-\beta T}] \ln \tilde{y}^*$ is supposed to be the same for all the economies and $u_{i;0,T}$ is the random disturbance of the data. Now we have a regression form of absolute convergence hypothesis.

5. CONVERGENCE CLUB COUNTRIES FOR ALBANIA

It is not a simple way to define a “convergence club” because of the heterogeneity of social aspect of Europe countries. Countries with millennium history and culture are significantly non similar between them. Even though, there are some characteristics which can make possible candidate countries for building “convergence club” for Albania. CEE and Western Balkans countries can construct a given “convergence club” for Albania. During 1989-1990 all these countries passed from a planned to a free market economy. After that, all of these countries have been evolved in the integration process in EU. Some of them are now part of EU realizing at least nominal convergence stipulated in the Treaty of Maastricht called nominal convergence. The fulfillment of these criteria’s is assessing the possibilities of sustainable economic growth under the conditions of promoting the public policies of achieving the convergence process in real term. Real convergence is defined as incomes convergence, prices convergence, productivity convergence, educational standards convergence, infrastructure development, economic and social cohesion, structural convergence with EU economies, etc. One of most important indicator of real convergence remains income per capita

Kornai [9] suggested different characteristics that make similar CEE countries during 1990-2004 in the convergence

TABLE 1
NOMINAL CONVERGENCE CRITERIA FOR EU MEMBER STATES

Indicators	Maastricht criteria
Inflation rate (% , annual average)	Under 1.5 pp over the average of the most successful 3 EU Member States
Long-term interest rate (% per year)	Under 2 pp over the average of the most successful 3 EU Member States
The exchange rate as compared to euro (maximum appreciation/depreciation in percentage compared to two years average***)	± 15 %
General government balance (% of GDP)	Under 3%
General government gross debt (% of GDP)	Under 60%

process. He called “great transformation” the process of transition of these countries because of: direction of market economy and democracy: parallel complete transformation in all spheres, the transformation was not violent and peaceful in sense of no use of military intervention, the incredible speed of transformation. Marelli and Signorelli [5] believe that the core of “great transformation” was the institutional change, which interacts with additional spheres: economic growth and development, structural change and economic performance, inequality and labor market, relations and shocks in global shocks in the global economy. These countries can be defined as a vanguard of “convergence club”.

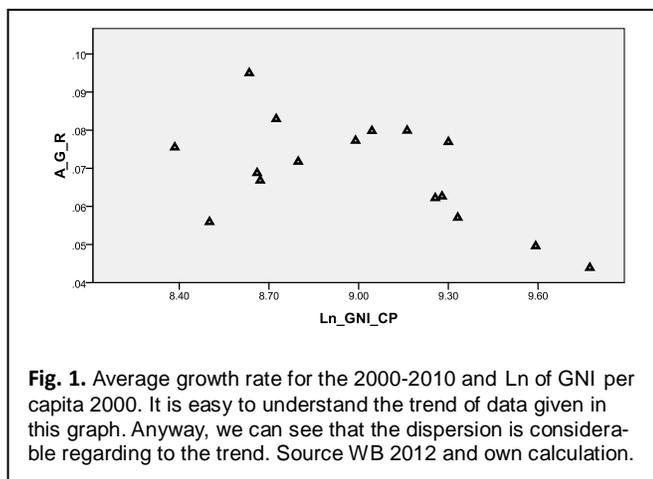
One of the most structural factors, which make possible the convergence, is the adoption *acquis communautaire*. Continued work in structural convergence along the lines of the *acquis communautaire* is crucial also in candidates like Macedonia, Montenegro and Croatia, and potential candidate countries like Albania, Bosnia Herzegovina, Serbia and Kosovo.

For Balkan countries the characteristics of convergence club are stronger. Central European Free Trade Agreement (CEFTA) have positive impacts in the regional integration process expanding trades and eliminating barriers between countries, providing protection of intellectual property rights, harmonizing policy and competition strategies facilitating sustainable growth and development in Balkan. This second group of countries completes the “convergence club” for Albania, which is evidently the last position in terms of Income per capita (PPP).

6. DATA ANALYSES OF ABSOLUTE CONVERGENCE

We are going to test for absolute beta convergence hypothesis using data for CEE and Western Balkans countries because of the “convergence club” characteristics. Log linear Regression between average of per capita income growth and log of initial income per capita is to be tested for absolute convergence hypothesis. The period observed in the cross-sectional data, due to particular conditions, is 2000-2010. The crisis of '97 in Albania and the war of Ex-Yugoslavian countries in 1999 may cause problems in the normal trend of data. By this facts, initial level of income per capita starts from 2000. In the data analyses is considered GNI per capita. In fact migration factor is present in different countries in the sample. Romania, Bulgaria, Estonia, Albania, Montenegro, Bosnia Herzegovina and others, showed a migratory movement towards western countries after 1990. GNI maybe reflect better the technological level, human and social capital and preferences embodied in national population. All this factors are considered significant in growth literature. Turning to the data analyses SPSS program is used to test the convergence hypothesis by OLS method and the regression is given by the equation

$$\sum_{t=0}^{11} g_t / T = b_0 - b_1 \ln y_{i,0} + u_{i,0,T} \quad (24)$$



Regression results are in support of the theoretically framework. It is found negative coefficient of Ln y₀ (b₁=-0.018) and it is statistically significant at 5%. The model can explain 30.2% of the variation in average of income per capita growth rate (R square =0.302) and residuals seems to have normal distribution. By calculation of the speed of convergence, the rate of which Gross national income per capita tends to steady state level, is around 2% per year by this calculation $-0.018 = -(1 - e^{-\beta \cdot 11})/11$. Barro and Sala-I-Martin suggested that the time for which Ln(y(t)) is in the halfway between Ln(y₀) and Ln(y*) is given by Lucas rule $70/\beta$. Lucas explain that considering y as a income per capita at time t and let y₀ be some initial value of per capita income. Then $y = y_0 e^{gt}$. The time it takes per capita income to double is given by the time t* at which $y = 2y_0$. Therefore, $2y_0 = y_0 e^{gt}$ implies $t^* = \ln 2/g$. Regarding to Barro and Sala-I-Martin (2004) is suggested that halfway convergence is the time for which equation

$\ln \tilde{y} = [1 - e^{-\beta t}] \ln \tilde{y}^* + e^{-\beta t} \ln \tilde{y}_0$ satisfies the condition $e^{-\beta t} = 1/2$. The halfway in time is therefore $[\ln(2)/\beta] = 0.69/0.02$. If we consider the result for $\beta = 0.02$ we can establish the time $t = 34.4$ years. This means that starting from the initial condition of income per capita only after 34 years will be in halfway of achieving the steady state level.

7. CONCLUSION

Absolute convergence hypothesis, implied by neoclassical model, consists in the process by which poor countries can thatch up rich ones. This hypothesis is not supported by empirical studies for the world. It is more acceptable conditional convergence hypothesis which means that counties with similar characteristics can converge in the term of income per capita creating “convergence club”. For Albania “convergence club” can be represented by CEE and Western Balkans countries because of, the similar characteristics in changing from commanded to free market economic system, facing to the European integration process in achieving nominal convergence postulated by Treaty of Maastricht and the adoption of *acquis communautaire*, the common economic agreement like CEFTA, etc. Considering cross-sectional data set for the period 200-2010 and using OLS method is estimated that speed of absolute convergence (β convergence) toward steady state level is around 2% per wear. Anyway, the model can explain 30 % of the variation in average of income per capita growth rate.

REFERENCES

- [1] A. Maddison, “Phases of Capitalist Development” *Oxford, Oxford University Press*, 1982.
- [2] D. Acemoglu, “Labor- and Capital-Augmenting Technical Change” <http://web.mit.edu/daron/www/lab-cap5.pdf>, Jan. 2000.
- [3] D. Cass, “Optimum growth in an aggregative model of capital accumulation,” *Review of Economic Studies*, pp.32, 233-240, 1965.
- [4] D. Raluca and C. Mirela, “Romania between real and nominal convergence,” *University of Craiova, Faculty of Economy and Business Administration*, 2007 <http://steconomice.uoradea.ro/anale/volume/2007/v2-finances-accounting-and-banks/62.pdf>.
- [5] E. Marelli and M. Signorelli, “Economic Growth and structural features of transition” pp. 11-56. PALGRAVE MACMILLAN Basingstoke, Hampshire RG21 6XS, 2010.
- [6] European Commission <http://www.cefta2006.com/>
- [7] F. P. Ramsey, “A Mathematical Theory of Saving”, *the Economic Journal*, Vol. 38, No. 152. (Dec), pp. 543-559, Dec. 1928.
- [8] G. Mankiw, D. Romer, D. Weil, “A Contribution to the Empirics of Economic Growth,” *The Quarterly Journal of Economics*, Volume 107, issue 2, pp. 407 - 437, May 1992.
- [9] J. Kornai, “The great transformation of Central Eastern Europe: Success and Disappointment,” *Economics of Transition*, pp. 2, 14, 207-244, Apr. 2006.
- [10] M. Abramovitz, “Catching Up, Forging Ahead and Falling Behind.” *Journal of Economic History*: 385-406. Jun. 1986.

- [11] P. Romer, "Endogenous Technological Change," *The Journal of Political Economy*, Vol. 98, No. 5, Part 2, pp. S71-S102, Oct.1990.
- [12] P. Romer, "Increasing Return and Long-Run Growth," *The Journal of Political Economy*, Vol. 94, Issue 5, pp. 1002 - 1037, Oct. 1986.
- [13] R. Barro and X. Sala-I-Martin, "Convergence", *Journal of Political Economy*, 100 (May): 223-252, May 1992.
- [14] R. Barro X. and Sala-I-Martin, "Economic Growth", 2nd edition, pp. 23-71, 462-496. The MIT Press Cambridge, Massachusetts, 2004.
- [15] R. E. Lucas, "On the mechanics of Economic Development," *Journal of Monetary Economics* 22, pp. 3-42, Feb. 1988.
- [16] R. M. Solow, "A Contribution to the Theory of Economic Growth," *the Quarterly Journal of Economics*, Vol. 70, No. 1, pp. 65-94, Feb. 1956.
- [17] R. M. Solow, "Technical Change and the Aggregate Production Function," *the Review of Economics and Statistics*, Vol. 39, No. 3, pp. 312-320, Aug. 1957.
- [18] R. Martin and A. Winkler, "Real Convergence in Central, Eastern and South-Eastern Europe" pp. 68-88, 124-152. PALGRAVE MACMILLAN Basingstoke, Hampshire RG21 6XS, 2009.
- [19] T. Koopmans, "On the concept of optimal economic growth," *Econometric Approach to Development Planning*, chap. 4, pp. 225-87, 1965.
- [20] T. Swan, "Economic Growth and Capital Accumulation," *Economic Record* 32.63, pp. 334-361, 1965.
- [21] W. J. Baumol, "Productivity Growth, Convergence, and Welfare: What the Long Run Data Show," *IU American Economic Review* 75, 1072-85, Aug. 1985.
- [22] World Bank <http://data.worldbank.org/topic/economic-policy-and-external-debt>
- [23] X. Sala-I-Martin, "The classical approach to convergence analysis" *Economics working paper* 117, Yale University and Universitat Pompeu Fabra, Jun. 1995.

ANNEX

TABLE 2

REGRESSION INPUT DATA

Country Name	Average GNI growth rate 2000-2010	GNI per capita 2010, PPP (current international \$)
Albania	0.075497	4380
Bosnia Herzegovina	0.055917	4920
Croatia	0.062631	10710
Czech Republic	0.049575	14650
Estonia	0.079887	9530
Hungary	0.057055	11290
Latvia	0.07725	8020
Lithuania	0.079804	8470
Macedonia, FYR	0.066803	5830
Montenegro	0.071743	6620
Romania	0.094989	5620
Slovenia	0.043893	17560
Slovak Republic	0.076913	10940
Poland	0.062195	10470
Serbia	0.068773	5770

Bulgaria	0.082936	6150
----------	----------	------

TABLE. 3

REGRESSION OUTPUTS

Unstandardized Coefficients		Standardized Coefficients	t	Sig.
B	Std. Error	Beta		
.234	.067		3.490	.004
-.018	.007	-.550	-2.464	.027

a. Dependent Variable: A_G_R

TABLE. 4

MODEL SUMMARY

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.550a	.302	.253	.01160

a. Predictors: (Constant), Ln_GNI_PC

b. Dependent Variable: Av_G_R

A Discrete-Event Coordinates Methodology for A Composite Curve

I. Akiije

ABSTRACT - This paper gives information about the results of investigation on a discrete-event simulation methodology to compute coordinates of a composite curve through Information and Communication Technology (ICT) system. Tangible modelling and simulation validity was done using general purpose application programs that is readily available rather than costly and tedious conventional hand method. In the methodology, deflection angle, tangential angles for the two transition curves and central circular arc, together with deviation angle computations, were successful modelled and simulated via ICT approach. Coordinates produced as the result of the process are useful set of numbers in Eastings and Northings to defining the entry transition curve, central circular arc and the exit transition curve of a highway or railroad composite curve at site. The conclusion in this study is that a discrete-event simulation is a suitable methodology for computing useful coordinates for setting out of a composite curve meant for a safe highway or railroad.

Index Terms – Coordinates, Composite, Discrete, Safe, Simulation, Tangential, Transition

1 INTRODUCTION

Discrete-event simulation could practically be done by hand calculations whereas for the amount of data that must be stored and manipulated for most real-world systems dictates that same is better done on a digital computer, Schriber and Brunner [1], Law and Kelton [2]. Discrete-event simulation goal in this paper is concerned with the modelling of a composite curve for highway or railroad in Information and Communication Technology (ICT) environment rather than traditional hand approach. According to Law and Kelton [2], discrete-event simulation concerns the modelling of a system by a representation in which the state variables change instantaneously at separate points. Simulation is a powerful tool for the evaluation and analysis of new system designs, modifications to existing systems and proposed changes to control systems and operating rules. Carson [3] claimed that conducting a valid simulation is both an art and a science. Adedimila and Akiije [4], Akiije [5], Akiije [6] considered simulation as the representation of physical systems and phenomena by computers, models and other equipment.

Simulation of coordinates for a composite curve by employing computer is the focus of this study. Coordinates are ordered set of numbers which specified the position or orientation of a point or geometric configuration relative to a set of axes in Eastings and Northings as claimed by Akiije [7].

According to Tiberius [8], McDonald [9], coordinates may be obtained from the record of some accurately located points of the framework of geodetic surveys or global positioning system (GPS).

Composite curve consists of entry transition curve, central circular arc and the exit transition curve. Types of transition curve in use for composite curve include clothoid or spiral, Bernoulli's Lemniscate, cubic parabola, cubical spiral and S-shaped. The need for transition curve paves in where a vehicle travelling on a straight course enters a curve of finite radius. At this juncture, Andrzej [10] claimed that such vehicle will be subjected to a sudden centrifugal force that will cause shock and sway. In order to avoid this centrifugal force, it is customary to provide a transition curve at the beginning and at the end of a circular curve.

A transition curve has a radius equal to infinity at the end of the straight and gradually reducing the radius to the radius of the circular curve where the later begins. Incidentally, the transition portion is useful for building up the centrifugal force gradually. Also, it provides a more aesthetically pleasing alignment and gradual application of the superelevation. Superelevation is the desirable raising of one edge of a roadway or a rail higher than the other along curve of a road or railway. The reason is to counteract centrifugal forces on passengers and vehicles or trains. The action brings comfort and safety to passengers and also prevents vehicles from overturning or sliding off the highway.

The objectives of this paper therefore is to provide a desirable automation platform for highway engineers to define coordinates as set of numbers in Eastings and Northings that are useful in an electronics environment for the design of a composite curve. Significantly with this study knowledge, computation of any segment of the composite

I. Akiije. Department of Civil and Environmental Engineering, University of Lagos, e-mail: akijieia@yahoo.com

curve can be modelled in a particular module and be simulated. Also, this study can be appreciated for where local computation modification is to be made in the modelling module; it is always made at ease. At this facet, by selection of an existing value and replacement of same by a new one, the required results are generated automatically. This technique is a powerful tool for geometric modelling and simulation of a composite curve coordinates. The justification for this study is that the use of digital computer is a most common computing device and a cheaper means of computation rather than conventional hand approach. In this study, digital computer performs operations on data represented in digital or number form with continuity and reusability capability for further necessary computation coordinates.

It is worthy of note that, the process of minimization and elimination of errors caused by instrumental imperfections or human operation by manual approach is a tedious work in order to accurately determine coordinates of points of a composite curve. However, GPS receiver can locate points in coordinates to an accuracy of ± 0.02 m as claimed by Uren and Price [11]. Therefore it is a useful instrument for defining coordinates points developed by the discrete-event methodology for a composite curve. GPS consists of three segments called the space segment, control segment and user segment. The use of GPS is advanced in this study at user segment level to capture coordinates of field stations and intersection points as data to compute coordinates for chainage points to define composite curve at the highway site. Also, Microsoft Excel a general purpose application program has been used to generate necessary chainage points coordinates for composite curve through computation via simulation modelling in an electronics office.

2 CONCEPTUAL FRAMEWORK

Transition curve chosen in this study is the cubic parabola simply because its formulae and calculations are easier to show in written form as claimed by Uren and Price [11]. Figure 1 is showing a typical framework of composite curve. The deflection angle is θ as measured on site or in the office by ICT methods rather than conventional hand method of using protractor. The radius R of the composite curve as required by the design speed V based upon design standards is possible by using Equation 1. The rate of change of radial acceleration c is defined by equation 2. Transition curve length $L_T = TT_1 = T_2U$ as defined by Equation 3.

In Figure 2, the shift at YG or WK is bisected by the transition curve and the transition curve is bisected by the shift. The shift S is defined by Equation 4. The tangent length IT = IU is defined by Equation 5. The maximum deviation

angle ϕ_{max} at the common tangent between the transition and the circular curve is defined by Equation 6. The length of the circular arc T_1T_2 is defined by Equation 7. The total length of the composite curve L_{total} is defined by Equation 8.

Figure 3 is showing the relationship between deviation angle ϕ and tangential angle δ . Deviation angle ϕ is defined by Equation 9. The formula showing the relationship between deviation angle ϕ and tangential angle δ is in Equation 10. The tangential angle δ for transitional curves is represented by Equations 11 and 12. The tangential angle α of the central circular arc is defined by Equations 13 and 14. Superelevation is calculated using Equation 15.

A technique of establishing a composite curve relies extensively on the use of transparent templates by conventional hand methodology. Pre-computed data from template are compiled into a design table that can be used as an input for the computation of the composite curve referenced completely to a coordinate system. The use of templates and design table relieves the design engineer of much computation and ensures more standardization and higher design standards. Although the conventional hand approach is a standard approach, it is tedious and it is not made in an electronic environment. In recent time, individual engineer's productivity has been increased by the use of automated procedures that rely extensively on electronic computer. Automated computation of coordinates to readily generate a composite curve is the focus of this study.

Design and setting out of the composite curve on site using coordinates methodology require the use of both rectangular and polar coordinates as claimed by Akijje [6]. The methodology requires the determination of eastings, northings and whole circular bearing of each chainage point along the alignment centre line as found in Equations 16 and 17. The variables for Equations 16 and 17 are in Figures 4 and 5. These variables include: E_A is Easting of A; E_B is Easting of B; N_A is Northing of A; N_B is Northing of B; ΔE_{AB} is eastings difference from A to B; ΔN_{AB} is northings difference from A to B; D_{AB} is horizontal length of AB; θ_{AB} is whole-circle bearing of line AB.

Hence, in this study, an innovative coordinates methodology of using GPS together with Microsoft Excel is introduced for appropriate cheaper technology of computing coordinates for designing and setting out of a composite curve at a site. In this new methodology, the use of horizontal control points supposedly established by surveyor near the alignment that may not be readily found is not necessary since

GPS could be used to capture coordinates of essential points that are required.

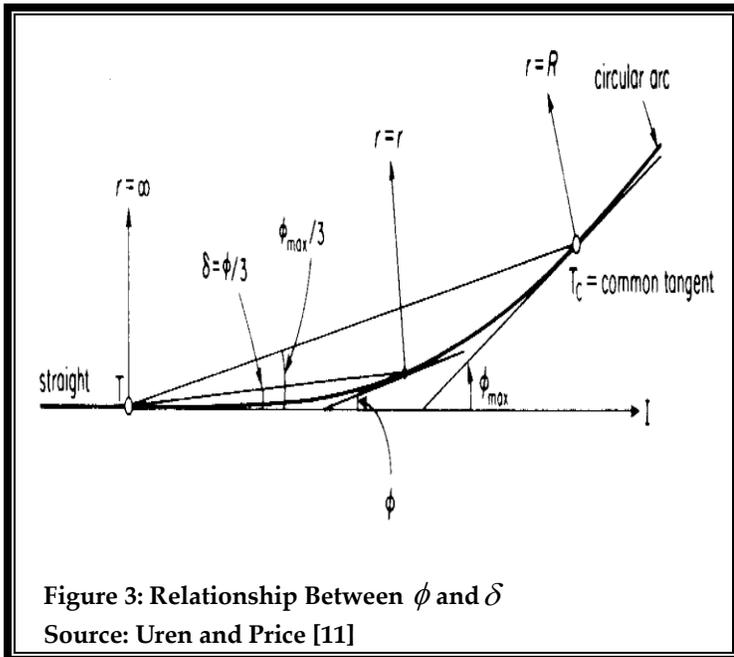
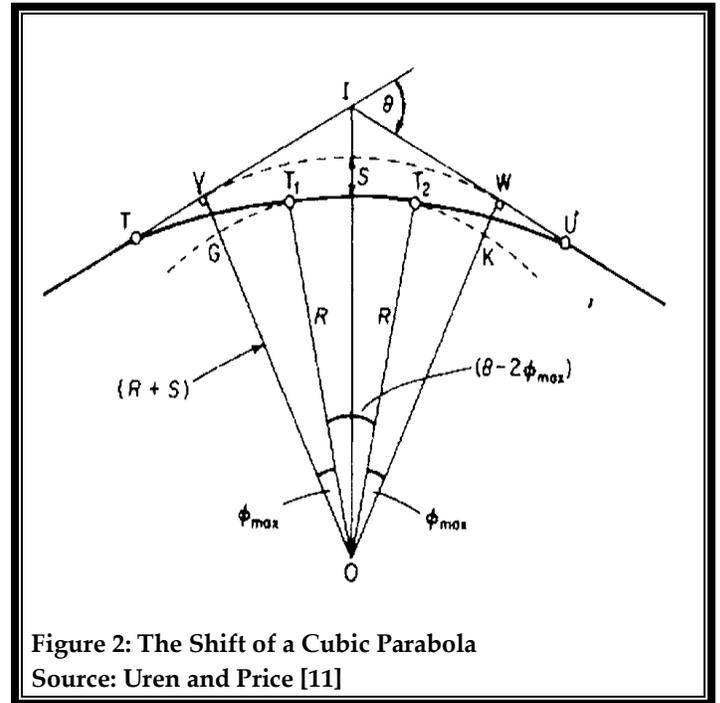
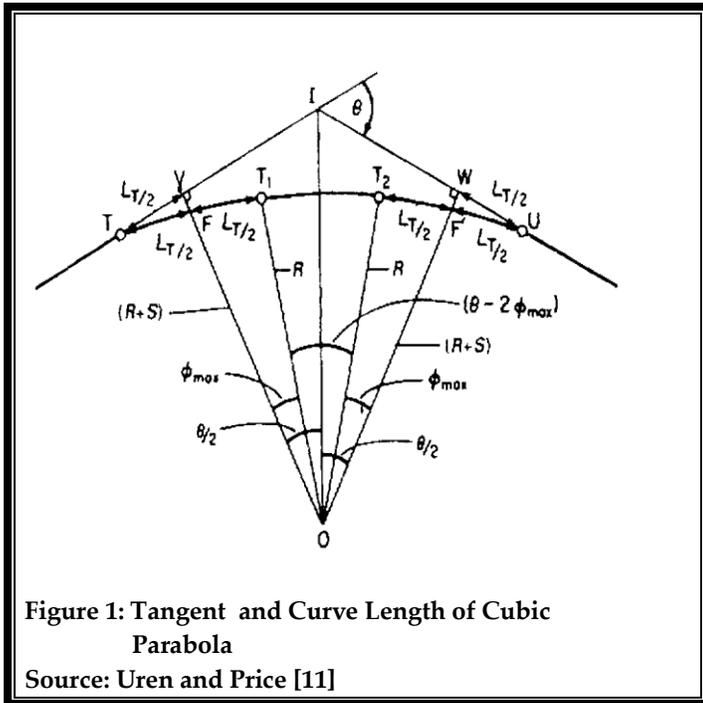
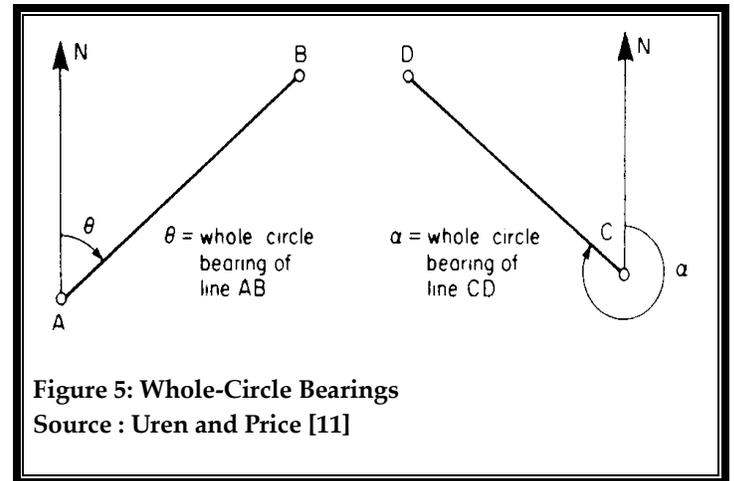
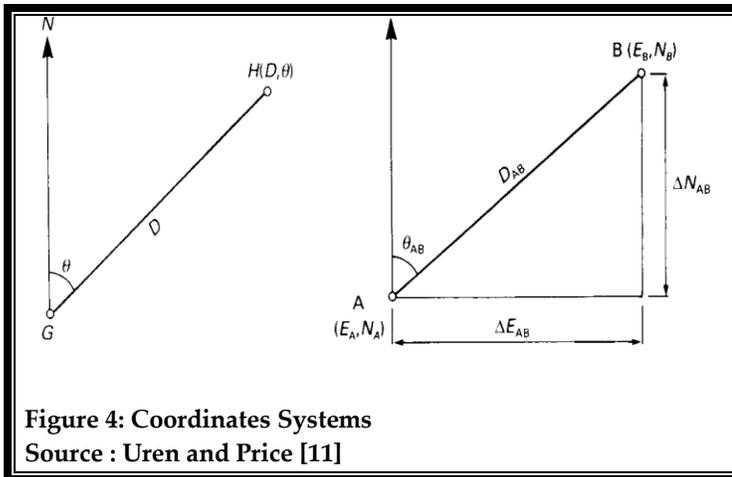


TABLE 1: VARIABLE EQUATIONS FOR A COMPOSITE CURVE

$\frac{V^2}{R} = 14.14$ for $V = 100$ km/h (1)	$c \leq 0.3$ m/s ² (2)
$L_T = \frac{V^3}{3.6^3 cR}$ (3)	$S = \frac{L_T^2}{24R}$ m (4)
$IT = IU = (R + S) \tan\left(\frac{\theta}{2}\right) + \frac{L_T}{2}$ m (5)	$\phi_{max} = \frac{L_T}{2R}$ radians (6)
$L_{circular\ arc} = T_1T_2 = R(\theta - 2\phi_{max})$ m (7)	$L_{total} = TT_1 + T_1T_2 + T_2U = L_T + R(\theta - 2\phi_{max}) + L_T$ m (8)
$\phi = \frac{l^2}{2RL_T}$ radians (9)	$\delta_{max} = \frac{\phi_{max}}{3}$ radians (10)
$\delta = \frac{l^2}{6RL_T}$ radians (11)	$\delta = \frac{l^2}{6RL_T} \times \frac{180}{\pi}$ degrees (12)
$\alpha = \frac{l}{R}$ radians (13)	$\alpha = \frac{180 \times l}{2\pi R}$ (14)
$e\phi\% = \frac{V^2}{2.828R}$ (15)	$E_B = E_A + \Delta E_{AB} = E_A + D_{AB} \sin \theta_{AB}$ (16)
$N_B = N_A + \Delta N_{AB} = N_A + D_{AB} \cos \theta_{AB}$ (17)	



3 MATERIALS AND METHODOLOGY

Data capturing and setting out materials for this study include GPS receiver. GPS receiver was able to connect with more than four satellites before coordinates readings and bookings. Computer workstation materials used to design the composite curve being study in this paper comprise the hardware and software components. The hardware components include systems board, central processing unit, memory, disks, a monitor, keyboard as an input device and printer as an output device. The software components used is the Microsoft Excel. Microsoft Excel is a spreadsheet with a table of cells with unique addresses for each cell. The important thing about the table is that data were entered into cells as labels and formulas written for the manipulation of same as modelling. Data modelled gave numerical results as simulation.

In this paper, a road leading to a fruit processing plant is used as a case study. A GPS receiver was used to capture coordinates (Table 2) of the beacons at the beginning (E_B, N_B) , at an intersection point (E_I, N_I) and at the point where the road to the factory stops (E_S, N_S) . The deflection angle of the connected two straights was determined within spreadsheet using Table 2. Based on the design speed of 85 km/h, the length of the transition curve and the tangent lengths were determined in Table 3. Also in Table 3, the determination of the through chainage of the beginning of entry transition curve T and the through chainage of the end of entry transition T_1 was defined. Table 4 is the modelling module of tangential angles for through chainages of the entry transition curve. Table 5 is the modelling module for the central circular arc of the composite curve. Table 6 is the modelling module of chord lengths tangential angles for through chainage of the central circular

curve. Table 7 is the modelling module for exit transition curve. Table 8 is the modelling module of chord lengths tangential angles for through chainage of the exit transition curve. Table 9 is the modelling module for coordinates of tangent point T , of entry transition curve, chainage 1537.088. Table 10 is the modelling module for coordinates of the initial sub-chord length for point C_1 , of the entry transition curve at chainage 1550.000.

Table 11 is the modelling module for coordinates first general chord length C_2 , of the entry transition curve at chainage 1575.000. Table 12 is the modelling module for coordinates the second general chord length C_3 , for the entry transition curve at chainage 1600.000. Table 13 is the modelling module for coordinates of the final sub-chord length for point T_1 , of the entry transition curve at chainage 1610.214. Table 14 is the modelling module for coordinates of initial sub-chord length end point C_4 , of circular arc, chainage 1625.000 m.

Table 15 is the modelling module for coordinates of the final sub-chord length end point T_2 , of the circular arc, chainage 1646.833 m. Table 16 is the modelling module for coordinates of tangent point U , of exit transition curve, chainage 1719.960 m. Table 17 is the modelling module for coordinates of the initial sub-chord length for point C_7 , of the exit transition curve, chainage 1700.000 m. Table 18 is the modelling module for coordinates first general chord length point C_6 , of exit transition curve, chainage 1675.000 m. Table 19 is the modelling module for coordinates for the second general chord length point C_5 , of exit transition curve, at chainage 1650.000 m.

TABLE 2: THE MODELLING MODULE FOR DEFLECTION ANGLE DETERMINATION

	C	D	E	F	G	H
2	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
3	$E_B =$	971695.890	m	Distance $D_{BI} =$ $(\Delta E_{BI}^2 + \Delta N_{BI}^2)^{0.5} =$	$=(D9^2 + D10^2)^{0.5}$	m
4	$N_B =$	678164.460	m	Distance $D_{BS} =$ $(\Delta E_{BS}^2 + \Delta N_{BS}^2)^{0.5} =$	$=(D11^2 + D12^2)^{0.5}$	m
5	$E_1 =$	972796.670	m	Bearing $\theta_{BI} = \tan^{-1}$ $(\Delta E_{BI} / \Delta N_{BI}) =$	$=ATAN(D9/D10)$	radians
6	$N_1 =$	679364.870	m	Bearing $\theta_{BS} =$	$=G5^*180/22^*7$	degrees
7	$E_5 =$	974123.000	m	Bearing $\theta_{BS} = \tan^{-1}$ $(\Delta E_{BS} / \Delta N_{BS}) =$	$=ATAN(D11/D12)$	radians
8	$N_5 =$	680364.300	m	Bearing $\theta_{BS} =$	$=G7^*180/22^*7$	degrees
9	$\Delta E_{BI} = E_1 - E_B =$	=D5-D3	m	Deflection angle $\theta =$ Bearing $\theta_{BI} -$ Bearing $\theta_{BS} =$	$=G7-G5$	radians
10	$\Delta N_{BI} = N_1 - N_B =$	=D6-D4	m	Deflection angle $\theta =$	$=G8-G6$	degrees
11	$\Delta E_{BS} = E_5 - E_1 =$	=D7-D5	m	Deflection angle $\theta =$	$10^{\circ} 29' 06''$	dms
12	$\Delta N_{BS} = N_5 - N_1 =$	=D8-D6	m			

TABLE 3: THE MODELLING MODULE FOR ENTRY TRANSITION CURVE

	C	D	E	F	G	H
19	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
20	Design speed, V =	85	Km/h	Change of Intersection point, CH.I =	=G3	
21	Rate of change of radial acceleration, c =	0.3		Through chainage of T = CH.I - IT =	=G20-D28	m
22	Superelevation, e% =	0.025	m/s ³	Through chainage of T ₁ = CH.I - IT + L _T =	=G21+D26	m
23	Side friction factor, f =	0.07		Through chainage multiple interval value =	25	m
24	Radius, R = $V^2 / (127(e\% + f)) =$	$=D20^2 / (127^*(D22 + D23)) =$	m	Next multiple of 25 m for chainage after T =	1550	m
25	Design Radius =	600	m	Last multiple of 25 m for chainage before T ₁ =	1600	m
26	Length of transition curve $L_T = V^3 / (3.6^3 c^2 R) =$	$=D20^3 / ((3.6^3 c^2 R) * D21 * D25) =$	m	Length of initial sub chord =	=G24-G21	m
27	Shift = $L_T^2 / 24R =$	$=D26^2 / 24 * D25 =$	m	Length of general chord =	25	m
28	Tangent length, IT = $(R + S) \tan \theta / 2 + LT / 2 =$	$=(D25 + D27) * TAN(G9/2) + D26/2 =$	m	Length of final sub-chord =	=G22-G25	m
29	Tangent length, IU = IT =	=D28	m			

TABLE 4: THE MODELLING MODULE OF TANGENTIAL ANGLES FOR THROUGH CHAINAGES OF THE ENTRY TRANSITION CURVE

	B	C	D	E	F	G
32	LABEL	MODELLING				LABEL
33	Chainage	Chainage Identification	Chord Length	Cumulative Length	Cumulative Clockwise Tangential Angle From T Relative To TL, $\Delta = L^2 / 6RL_T$	Tangential Angles Identification
34						
35	(m)		(m)	(m)	radians	
36	=G21	T	0	0	=E36^2 / (6^*SDS25^*SDS26)	
37	=B36+D37	C ₁	=G26	=D37	=E37^2 / (6^*SDS25^*SDS26)	δ_1
38	=B37+D38	C ₂	=G27	=E37+D38	=E38^2 / (6^*SDS25^*SDS26)	δ_2
39	=B38+D39	C ₃	=G27	=E38+D39	=E39^2 / (6^*SDS25^*SDS26)	δ_3
40	=B39+D40	T ₁	=G28	=E39+D40	=E40^2 / (6^*SDS25^*SDS26)	δ_{max}
41	SUM =		=SUM(D36:D40)	(checks)	=D26/2/D25/3	$\phi_{max}/3 = L_T/6R/3$

TABLE 5: THE MODELLING MODULE FOR THE CENTRAL CIRCULAR ARC

	C	D	E	F	G	H
45	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
46	Maximum Deviation Angle $\phi_{max} = L_T / (2R) =$	$=D26 / (2 * D25) =$	radians	Last multiple of 25 m for chainage before T ₁ =	1625	m
47	Length of circular arc $L_c = R(\theta - 2\phi_{max}) =$	$=D25^2 * (G9 - 2 * D46) =$	m	Length of initial sub chord =	=D49-G22	m
48	Through chainage of T ₁ = CH.T ₁ + L _C =	=G22+D47	m	Length of general chord =	0	m
49	Next multiple of 25 m for chainage after T ₁ =	1625	m	Length of final sub-chord =	=D48-G46	m

TABLE 6: THE MODELLING MODULE OF CHORD LENGTHS TANGENTIAL ANGLES FOR THROUGH CHAINAGE OF THE CENTRAL CIRCULAR CURVE

	B	C	D	E	F	G
52	LABEL		MODELLING			LABEL
53	Chainage	Chainage Identification	Chord Length	Tangential Angle for Each Chord $\alpha = 1/2R$	Cumulative Clockwise Tangential Angle From T Relative to the Common Tangent	Tangential Angles Identification
54						
55	(m)		(m)	radians	radians	
56	=G22	T ₁	0	=D56/2/SDS25	=E56	
57	=B56+D57	C ₄	=G47	=D57/2/SDS25	=F56+E57	α_1
58	=B57+D58	T ₂	=G49	=D58/2/SDS25	=F57+E58	α_2
59	SUM =		=SUM(D56:D58)	(checks)	=D47/D25/2	$(\theta - 2\phi_{max})/2$

TABLE 7: THE MODELLING MODULE FOR EXIT TRANSITION CURVE

	C	D	E	F	G	H
63	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
64	Through Chainage of T ₂ =	=D48	m	Last Multiple of 25m for Chainage Before U =	1700	m
65	Length of Exit Transition Curve =	=D26		Initial Sub-Chord Length from U =	=D66-G64	m
66	Through Chainage of U =	=SUM(D64:D65)		Length of General Chord =	25	m
67	First Multiple of 25m for Chainage After T ₂ =	1650	m	Final Sub-Chord Length to T ₂ =	=D67-D64	m

TABLE 8: THE MODELLING MODULE OF CHORD LENGTHS TANGENTIAL ANGLES FOR THROUGH CHAINAGE OF THE EXIT TRANSITION CURVE

	B	C	D	E	F	G
70	LABEL		MODELLING			LABEL
71	Chainage	Chainage Identification	Chord Length	Cumulative Length	Cumulative Clockwise Tangential Angle From U Relative To UI, $\Delta = L^2/6RLT$	Tangential Angles Identification
72						
73	(m)		(m)	(m)	radians	
74	=D66	U	0	0	=E74^2/(6*SDS25*SDS26)	
75	=B74-D75	C ₇	=G65	=D75	=E75^2/(6*SDS25*SDS26)	δ_7
76	=B75-D76	C ₆	=G66	=E75+D76	=E76^2/(6*SDS25*SDS26)	δ_6
77	=B76-D77	C ₅	=G66	=E76+D77	=E77^2/(6*SDS25*SDS26)	δ_5
78	=B77-D78	T ₂	=G67	=E77+D78	=E78^2/(6*SDS25*SDS26)	δ_{max}
79	SUM =		=SUM(D74:D78)	(checks)	=F41	$\phi_{max}/3 = L_T/6R/3$

TABLE 9: THE MODELLING MODULE FOR COORDINATES TANGENT POINT T, OF ENTRY TRANSITION CURVE, CHAINAGE 1537.088 M

	C	D	E	F	G	H
83	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
84	Length of T ₁ =	=D29	m	Cos Bearing of T ₁ =	=COS(D87)	radians
85	E ₁ =	=D5	m	$\Delta E_{T1} = D_{T1} \times \sin$ Bearing of T ₁ =	=D84+D88	m
86	N ₁ =	=D6	m	$\Delta N_{T1} = D_{T1} \times \cos$ Bearing of T ₁ =	=D84+G84	m
87	Bearing of T ₁ =	=G5	m	E _T = E ₁ - ΔE_{T1}	=D85-G85	m
88	Sin Bearing of T ₁ =	=SIN(D87)	radians	N _T = N ₁ - ΔN_{T1}	=D86-G86	m

TABLE 10: THE MODELLING MODULE FOR COORDINATES INITIAL SUB-CHORD LENGTH C₁, OF ENTRY TRANSITION CURVE, CHAINAGE 1550.000 M

	C	D	E	F	G	H
93	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
94	Length of TC ₁ =	=D37	m	$\Delta N_{T1} =$	=D97-D99	m
95	$\delta_1 =$	=F37	radians	Bearing of T ₁ = arc Tan ($\Delta E_{T1}/\Delta N_{T1}$) =	=ATAN(D100/G94)	radians
96	E ₁ =	=D85	m	Bearing of TC ₁ = Bearing of T ₁ + $\delta_1 =$	=G95+D95	radians
97	N ₁ =	=D86	m	$\Delta E_{TC1} = TC_1 \sin$ bearing of TC ₁ =	=D94+SIN(G96)	m
98	E _T =	=G87	m	$\Delta N_{TC1} = TC_1 \cos$ bearing of TC ₁ =	=D94+COS(G96)	m
99	N _T =	=G88	m	E _{C1} = E _T + ΔE_{TC1}	=D98+G97	m
100	$\Delta E_{T1} =$	=D96-D98	m	N _{C1} = N _T + ΔN_{TC1}	=D99+G98	m

TABLE 11: THE MODELLING MODULE FOR COORDINATES FIRST GENERAL CHORD LENGTH C₂, OF ENTRY TRANSITION CURVE, CHAINAGE 1575.000 M

	C	D	E	F	G	H
104	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
105	LENGTH OF TC ₁ =	=D37	m	bearing C ₁ C ₂ = bearing TC ₁ + Y =	=G96+D111	radians
106	LENGTH OF C ₁ C ₂ =	=D38	m	$\Delta E_{C1C2} = C_1 C_2 \sin$ BEARING OF C ₁ C ₂ =	=D106+SIN(G105)	m
107	$\delta_1 =$	=F37	radians	$\Delta N_{C1C2} = C_1 C_2 \cos$ BEARING OF C ₁ C ₂ =	=D106+COS(G105)	m
108	$\delta_2 =$	=F38	radians	E _{C1} =	=G99	m
109	$\sin \beta_1 = (TC_1/C_1 C_2) \times \sin(\delta_2 - \delta_1) =$	=(D105/D106)*sin (D108-D107)	radians	N _{C1} =	=G100	m
110	$\beta_1 =$	=ASIN(D109)	radians	E _{C2} = E _{C1} + $\Delta E_{C1C2} =$	=G108+G106	m
111	Y = $\beta_1 + (\delta_2 - \delta_1) =$	=D110+D108-D107	radians	N _{C2} = N _{C1} + $\Delta N_{C1C2} =$	=G109+G107	m

TABLE 12: THE MODELLING MODULE FOR COORDINATES SECOND GENERAL CHORD LENGTH C_3 , OF ENTRY TRANSITION CURVE, CHAINAGE 1600.000 M

	C	D	E	F	G	H
115	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
116	Length of $TC_1 =$	=D37	m	$Y = \beta_2 + (\delta_3 - \delta_2) =$	=D123+D121-D120	radians
117	Length of $C_1C_2 =$	=D38	m	bearing $C_2C_3 =$ bearing $C_1C_2 + Y =$	=G105+G116	radians
118	Length of $TC_2 =$ $TC_1 + C_1C_2 =$	=D116+D117	m	$\Delta E_{C2C3} = C_2C_3 \sin$ bearing of $C_2C_3 =$	=D119*Sin(G117)	radians
119	Length of $C_2C_3 =$	=D39	m	$\Delta N_{C2C3} = C_2C_3 \cos$ bearing of $C_2C_3 =$	=D119*Cos(G117)	m
120	$\delta_2 =$	=F38	radians	$E_{C2} =$	=G110	m
121	$\delta_3 =$	=F39	radians	$N_{C2} =$	=G111	m
122	$\sin \beta_2 =$ $(TC_2/C_2C_3) * \sin(\delta_3 - \delta_2) =$	= (D118/D119) * SIN(D121-D120)	radians	$E_{C3} = E_{C2} + \Delta E_{C2C3} =$	=G120+G118	m
123	$\beta_2 =$	=ASIN(D122)	radians	$N_{C3} = N_{C2} + \Delta N_{C2C3} =$	=G121+G119	m

TABLE 13: THE MODELLING MODULE FOR COORDINATES OF THE FINAL SUB-CHORD LENGTH FOR POINT T_1 , CHAINAGE 1610.214 M

	C	D	E	F	G	H
127	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
128	Length of $TC_2 =$	=D118	m	$Y = \beta_3 + (\delta_{max} - \delta_3) =$	=D135+(D133-D132)	radians
129	Length of $C_2C_3 =$	=D39	m	bearing $C_3T_1 =$ bearing $C_2C_3 + Y =$	=G117+G128	radians
130	Length of $TC_3 =$ $TC_2 + C_2C_3 =$	=D128+D129	m	$\Delta E_{C3T1} = C_3T_1 \sin$ bearing of $C_3T_1 =$	=D131*Sin(G129)	m
131	Length of $C_3T_1 =$	=D40	m	$\Delta N_{C3T1} = C_3T_1 \cos$ bearing of $C_3T_1 =$	=D131*Cos(G129)	m
132	$\delta_3 =$	=F39	radians	$E_{C3} =$	=G122	m
133	$\delta_{max} =$	=F40	radians	$N_{C3} =$	=G123	m
134	$\sin \beta_3 = (TC_3/C_3T_1) * \sin(\delta_{max} - \delta_3) =$	= (D130/D131) * SIN(D133-D132)	radians	$E_{T1} = E_{C3} + \Delta E_{C3T1} =$	=G132+G130	m
135	$\beta_3 =$	=ASIN(D134)	radians	$N_{T1} = N_{C3} + \Delta N_{C3T1} =$	=G133+G131	m

TABLE 14: THE MODELLING MODULE FOR COORDINATES INITIAL SUB-CHORD LENGTH END POINT C_4 , OF CIRCULAR ARC, CHAINAGE 1625.000 M

	C	D	E	F	G	H
139	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
140	$\phi_{max} =$	=D46	radians	$\Delta E_{TC4} = \text{Length}$ $T_1C_4 \times \sin$ Bearing $T_1C_4 =$	=D145*Sin(D144)	m
141	Bearing $TC_4 =$	=G95	radians	$\Delta N_{TC4} = \text{Length}$ $T_1C_4 \times \cos$ Bearing $T_1C_4 =$	=D145*Cos(D144)	m
142	Bearing $T_1Z =$ Bearing $TC_4 + \phi_{max} =$	=SUM(D140:D141)	radians	$E_{T1} =$	=G134	m
143	$\alpha_1 =$	=F57	radians	$N_{T1} =$	=G135	m
144	Bearing $T_1C_4 =$ Bearing $T_1Z + \alpha_1 =$	=SUM(D142:D143)	radians	$E_{C4} = E_{T1} +$ $\Delta E_{TC4} =$	=G142+G140	m
145	Length $T_1C_4 =$	=D57	m	$N_{C4} = N_{T1} +$ $\Delta N_{TC4} =$	=G143+G141	m

TABLE 15: THE MODELLING MODULE FOR COORDINATES OF THE FINAL SUB-CHORD LENGTH END POINT T_2 , OF THE CIRCULAR ARC, CHAINAGE 1646.833 M

	C	D	E	F	G	H
149	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
150	$\alpha_1 =$	=F57	radians	$\Delta E_{C4T2} = \text{Length}$ $C_4T_2 \times \sin$ Bearing $C_4T_2 =$	=D155*Sin(D154)	m
151	$\alpha_2 =$	=F58	radians	$\Delta N_{C4T2} = \text{Length}$ $C_4T_2 \times \cos$ Bearing $C_4T_2 =$	=D155*Cos(D154)	m
152	$\lambda = (\alpha_1 + \alpha_2) =$	=SUM(D150:D151)	radians	$E_{C4} =$	=G144	m
153	Bearing $T_1C_4 =$	=D144	radians	$N_{C4} =$	=G145	m
154	Bearing $C_4T_2 =$ Bearing $T_1C_4 + \lambda =$	=SUM(D152:D153)	radians	$E_{T2} = E_{C4} + \Delta E_{C4T2} =$	=G152+G150	m
155	Length $C_4T_2 =$	=D58	m	$N_{T2} = N_{C4} + \Delta N_{C4T2} =$	=G153+G151	m

TABLE 16: THE MODELLING MODULE FOR COORDINATES TANGENT POINT U, OF EXIT TRANSITION CURVE, CHAINAGE 1719.960 M

	C	D	E	F	G	H
159	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
160	$\theta =$	$=G9$	radians	$\Delta N_{IU} = \text{Length IU} \times \text{Cos Bearing IU} =$	$=D163 * \text{COS}(D162)$	m
161	Bearing IU =	$=G95$	radians	$E_I =$	$=D5$	m
162	Bearing IU + Bearing IU + $\theta =$	$=\text{SUM}(D160:D161)$	radians	$N_I =$	$=D6$	m
163	Tangent length IU =	$=D29$	m	$E_U = E_I + \Delta E_{IU} =$	$=G161 + D164$	m
164	$\Delta E_{IU} = \text{Length IU} \times \text{Sin Bearing IU} =$	$=D163 * \text{SIN}(D162)$	m	$N_U = N_I + \Delta N_{IU} =$	$=G162 + G160$	m

TABLE 17: THE MODELLING MODULE FOR COORDINATES OF THE INITIAL SUB- CHORD LENGTH FOR POINT C₇, OF THE EXIT TRANSITION CURVE, CHAINAGE 1700.000 M

	C	D	E	F	G	H
168	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
169	Length of UC ₇ =	$=D75$	m	$\Delta N_U =$	$=D172 - D174$	m
170	$\delta_{7z} =$	$=F75$	radians	Bearing of UI = Arc Tan ($\Delta E_U / \Delta N_U$) =	$=\text{ATAN}(D175/G169)$	radians
171	$E_U =$	$=G163$	m	Bearing of UC ₇ = bearing of UI - $\delta_{7z} =$	$=G170 - D170$	radians
172	$N_U =$	$=G164$	m	$\Delta E_{UC7} = UC_7 \text{ Sin bearing of } UC_7 =$	$=D169 * \text{SIN}(G171)$	m
173	$E_I =$	$=D5$	m	$\Delta N_{UC7} = UC_7 \text{ Cos bearing of } UC_7 =$	$=D169 * \text{COS}(G171)$	m
174	$N_I =$	$=D6$	m	$E_{C7} = E_U - \Delta E_{UC7} =$	$=D171 - G172$	m
175	$\Delta E_U =$	$=D171 - D173$	m	$N_{C7} = N_U - \Delta N_{UC7} =$	$=D172 - G173$	m

4 RESULTS AND DISCUSSION

The summary of the modelling that processed the coordinates of points to define the composite curve is shown in Table 20. Table 21 is showing the summary of the simulation of coordinates for the design and setting out of the composite curve on site. Modelling modules of Table 2 through Table 19 allow iteration process by engineers to automate design variables through a discrete-event simulation design process to develop coordinates for a composite curve. This is accomplished through an integrated ICT design environment that links design activities of entry transition curve, circular arc and exit transition curve to produce coordinates for a composite curve.

TABLE 18: THE MODELLING MODULE FOR COORDINATES FIRST GENERAL CHORD LENGTH POINT C₆, OF EXIT TRANSITION CURVE, CHAINAGE 1675.000 M

	C	D	E	F	G	H
179	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
180	Length of UC ₇ =	$=D169$	m	bearing C ₇ C ₆ = bearing UC ₇ - Y =	$=G171 - D186$	m
181	Length of C ₇ C ₆ =	25	m	$\Delta E_{C7C6} = C_7C_6 \text{ Sin bearing of } C_7C_6 =$	$=D181 * \text{SIN}(G180)$	m
182	$\delta_{7z} =$	$=F75$	radians	$\Delta N_{C7C6} = C_7C_6 \text{ Cos bearing of } C_7C_6 =$	$=D181 * \text{COS}(G180)$	m
183	$\delta_6 =$	$=F76$	radians	$E_{C7} =$	$=G174$	m
184	$\text{Sin} \beta_7 = (UC_7 / C_7C_6) * \text{Sin}(\delta_6 - \delta_7) =$	$= (D180 / D181) * \text{Sin}(D183 - D182)$	radians	$N_{C7} =$	$=G175$	m
185	$\beta_7 =$	$=\text{ASIN}(D184)$	radians	$E_{C6} = E_{C7} - \Delta E_{C7C6} =$	$=G183 - G181$	m
186	$Y = \beta_7 + (\delta_6 - \delta_7) =$	$=D185 + D183 - D182$	radians	$N_{C6} = N_{C7} - \Delta N_{C7C6} =$	$=G184 - G182$	m

TABLE 19: THE MODELLING MODULE FOR COORDINATES FOR THE SECOND GENERAL CHORD LENGTH POINT C₅, OF EXIT TRANSITION CURVE, CHAINAGE 1650.000 M

	C	D	E	F	G	H
190	LABEL	MODELLING	UNIT	LABEL	MODELLING	UNIT
191	Length of UC ₇ =	$=D180$	m	$Y = \beta_6 + (\delta_5 - \delta_6) =$	$=D198 + D196 - D195$	radians
192	Length of C ₇ C ₆ =	$=D181$	m	bearing C ₆ C ₅ = bearing C ₇ C ₆ - Y =	$=G180 - G191$	radians
193	Length of UC ₆ = UC ₇ + C ₇ C ₆ =	$=D191 + D192$	m	$\Delta E_{C6C5} = C_6C_5 \text{ Sin bearing of } C_6C_5 =$	$=D194 * \text{Sin}(G192)$	radians
194	Length of C ₆ C ₅ =	25	m	$\Delta N_{C6C5} = C_6C_5 \text{ COS bearing of } C_6C_5 =$	$=D194 * \text{COS}(G192)$	m
195	$\delta_{6z} =$	$=F76$	radians	$E_{C6} =$	$=G185$	m
196	$\delta_{5z} =$	$=F77$	radians	$N_{C5} =$	$=G186$	m
197	$\text{Sin} \beta_6 = (UC_6 / C_6C_5) * \text{Sin}(\delta_5 - \delta_6) =$	$= (D193 / D194) * \text{Sin}(D196 - D195)$	radians	$E_{C5} = E_{C6} - \Delta E_{C6C5} =$	$=G195 - G193$	m
198	$\beta_6 =$	$=\text{ASIN}(D197)$	radians	$N_{C5} = N_{C6} - \Delta N_{C6C5} =$	$=G196 - G194$	m

Numerous design iterations for the purpose of improving and refining the coordinates to develop a composite curve without expending a large amount of time or effort are possible using the modelling tables. The developed modelling tables within spreadsheet environment are valuable features with the ability to view the resulting effect of the computation of coordinates and modifications. Hence, coordinates for a composite curve are automatically carried out via a discrete-event methodology without the need to conduct the numerous intermediate steps that have been associated with the conventional manual design method. The method developed here does not need the

involvement of vendors to complement design activities as in the case of specific purpose application programs and the system operation is not under license.

5 CONCLUSIONS AND RECOMMENDATIONS

The following conclusions and recommendations are derived from the investigation carried out in this study.

5.1 CONCLUSIONS

1. A new methodology of computing coordinates for designing a composite curve via a discrete-event simulation approach in an ICT environment has been vividly carried successfully.
2. This novel methodology has been successfully carried out by making use of coordinates generated by GPS on site and advancing same in Microsoft Excel to compute coordinate to design a composite curve.
3. The process allows coordinates to be computed for a composite curve by using various related variables that change instantaneously at separate points in different modelling module tables while making use of a readily available general purpose application program.

4. The result is similar to the conventional hand method of computation but this novel methodology is in an electronics environment.
5. The methodology is amenable to intranet and internet and allows various engineers to work on a composite curve for improvement optimally.
6. The process has the ability to enhance the productivity of highway engineers to conduct numerous iteration of computations for a composite curve coordinates for the purpose of improving and refining without expending a large amount of time or effort.

5.2 RECOMMENDATIONS

1. The methodology is highly recommended as a better alternative to the use of programming software, specific purpose application programs and conventional hand method in the computation of coordinates for design of road or rail composite curve.
2. The methodology is also highly recommended as a better alternative while setting out of road or rail composite curve.

TABLE 20: MODELLING MODULE SUMMARY OF THE COMPOSITE CURVE COORDINATES

	J	K	L	M
215	LABEL		MODELLING	
216	Point	Through Chainage	Coordinates	
217		m	mE	mN
218	T	1537.088	=G87	=G88
219	C ₁	1550	=G99	=G100
220	C ₂	1575	=G110	=G111
221	C ₃	1600	=G122	=G123
222	T ₁	1610.214	=G134	=G135
223	C ₄	1625	=G144	=G145
224	T ₂	1646.833	=G154	=G155
225	C ₅	1650	=G197	=G198
226	C ₆	1675	=G185	=G186
227	C ₇	1700	=G174	=G175
228	U	1719.96	=G163	=G164

TABLE 21: SUMMARY OF THE COMPOSITE CURVE COORDINATES SIMULATION MODULE

	J	K	L	M
215	LABEL		SIMULATION	
216	Point	Through Chainage	Coordinates	
217		m	mE	mN
218	T	1537.088	972734.745	679297.340
219	C ₁	1550.000	972743.478	679306.852
220	C ₂	1575.000	972760.521	679325.143
221	C ₃	1600.000	972777.999	679343.017
222	T ₁	1610.214	972785.411	679350.046
223	C ₄	1625.000	972796.175	679360.182
224	T ₂	1646.833	972812.696	679374.457
225	C ₅	1650.000	972814.848	679376.786
226	C ₆	1675.000	972834.147	679392.678
227	C ₇	1700.000	972853.922	679407.973
228	U	1719.960	972869.845	679420.009

REFERENCES

- [1] T. J. Schriber and D. T. "Brunner Inside Discrete-Event Simulation Software How It Works and Why It Matters," Proceedings of the 37th Conference on Winter Simulation, pp.167-177, 2005.
- [2] A. M. Law and W. D. Kelton. "Simulation Modeling and Analysis" 3rd Edition, Tata McGraw-Hill, New Delhi, 2007.
- [3] J. S. Carson. "Introduction to Modelling and Simulation," Proceedings of the 37th conference on winter simulation, pp.16-23, 2005.
- [4] A. S. Adedimila and I. Akijje. "National Capacity Development in Highway Geometric Design," Proceedings, National Engineering Conference. The Nigerian Society of Engineers, Abeokuta, 228-235, 2006.
- [5] I. Akijje and S. A. Adedimila. "A New Approach to the Geometric Design of Highway Alignments Using a Microcomputer," Proceedings of the International Conference in Engineering, paper 43, at the University of Lagos, Lagos, Nigeria, 2005.
- [6] I. Akijje. "An Optimal Control for Highway Horizontal Circular Curve Automation," Proceedings of International Conference on Innovations in Engineering and Technology, Faculty of Engineering, University of Lagos, Akoka, Lagos, pp 669-681, 2011.
- [7] I. Akijje. "An Analytical and Autograph Method for Highway Geometric Design," Ph.D. Thesis, School of Postgraduate Studies, University of Lagos, Akoka, Lagos, 2007.
- [8] C. Tiberius. "Standard Positioning Service," Handheld GPS Receiver Accuracy in GPS World, Vol. 14, No. 2 pp. 44-51, 2003.
- [9] K. McDonald. "GPS Modernization Global: Positioning System Planned Improvement," Proceedings ION Annual Meeting pp. 36-39, 2005.
- [10] Andrzej Kobryn. "Polynomial Solutions of Transaction Curves," Technical Papers, Journal of Surveying Engineering / Volume 137/Issue 3, 2011.
- [11] J. Uren and W. F. Price. "Surveying for Engineers." 5th Edition, Palgrave Macmillan, Great Britain, 2010.

S_β -compact and S_β -closed spaces

Alias B. Khalaf* and Nehmat K. Ahmed**

Abstract - The objective of this paper is to obtain the properties of S_β -compact and S_β -closed spaces by using nets, filterbase and S_β -complete accumulation points.

Index Terms— S_β -open sets, semi open sets, S_β -compact spaces, S_β -closed spaces, S_β -complete accumulation points.

1 INTRODUCTION AND PRELIMINARIES

It is well-known that the effects of the investigation of properties of closed bounded intervals of real numbers, spaces of continuous functions and solution to differential equations are the possible motivations for the formation of the notion of compactness. Compactness is now one of the most important, useful, and fundamental notions of not only general topology but also of other advanced branches of mathematics. Recently Khalaf A.B. et al. [13] introduced and investigated the concepts of S_β -space and S_β -continuity. A semi open subset A of a topological space (X, τ) is said to be S_β -open if for each $x \in X$ there exists a β -closed set F such that $x \in F \subseteq A$. The aim of this paper to give some characterizations of S_β -compact spaces in terms of nets and filter-bases. We also introduce the notion of S_β -complete accumulation points by which we give some characterizations of S_β -compact spaces. Throughout the present paper, (X, τ) and (Y, ν) or simply X and Y denote topological space. In [14] Levine initiated semi open sets and their properties.

Mathematicians gives in several papers interesting and different new types of sets. In [1] Abd-El-Moonsef defined the class of β -open set. In [18] Shareef introduced a new class of semi-open sets called S_p -open sets. We recall the following definitions and characterizations. The closure (resp., interior) of a subset A of X is denoted by clA (resp., $intA$). A subset A

of X is said to be semi-open [14] (resp., pre-open [15], α -open [16], β -open [1] regular open [19] and regular β -open [22]) set if $A \subseteq cl\ int A$, (resp., $A \subseteq int\ cl A$, $A \subseteq int\ cl\ int A$, $A \subseteq cl\ int\ cl A$, $A = int\ cl A$ and $A = \beta\ int\ \beta\ cl A$). The complement of S_β -open (resp., semi-open, pre-open, α -open, β -open, regular open, regular β -open) set is said to be S_β -closed (resp., semi-closed, pre-closed, α -closed, β -closed, regular closed, regular β -closed). The intersection of all S_β -closed (resp., semi-closed, pre-closed, β -closed) sets of X containing a subset A is called the S_β -closure (resp., semi-closure, pre-closure β -closure) of A and denoted by $S_\beta\ cl A$ (resp., $sclA$, $pclA$, $\beta\ cl A$). The union of all S_β -open (resp., semi-open, pre-open, β -open) set of X contained in A is called the S_β -interior (resp., semi-interior, pre-interior, β -interior) of A and denoted by $S_\beta\ int A$ (resp., $sintA$, $pintA$, $\beta\ int A$). The family of all S_β -open (resp., semi-open, pre-open, α -open, β -open, regular β -open, regular open, S_β -closed, semi-closed, pre-closed, α -closed, β -closed, regular β -closed, and regular closed) subset of a topological space X is denoted by $S_\beta\ O(X)$ (resp., $SO(X)$, $PO(X)$, $\alpha\ O(X)$, $\beta\ O(X)$, $R\ \beta\ O(X)$, $RO(X)$, $S_\beta\ C(X)$, $SC(X)$, $PC(X)$, $\alpha\ C(X)$, $\beta\ C(X)$, $R\ \beta\ C(X)$ and $RC(X)$). A subset A of X is called δ -open [21] if for each $x \in A$, there exists an open set B such that $x \in B \subseteq int\ cl B \subseteq A$. A subset A of a space X is called θ -semi-open [12] (resp., semi- θ -open [7] if for each $x \in A$, there exists a semi-open set B such that $x \in B \subseteq cl B \subseteq A$ (resp., $x \in B \subseteq S\ cl B \subseteq A$).

Definition 1.1 [15]. A topological space (X, τ) is said to be :

- 1- Extremely disconnected if $clV \in \tau$ for every $V \in \tau$.
- 2- Locally indiscrete, if every open subset of X is closed.

*Department of Mathematics, Faculty of Science, University of Duhok, Duhok City, Kurdistan- Region, Iraq.
E-mail: aliasbkhalf@gmail.com

** Co-Author is currently pursuing Ph. D. degree program in Mathematics at the Department of Mathematics, College of Education, Salahaddin University.
E-mail: nehmatbalin@yahoo.com

3- Hyper-connected if every non-empty open subset of X is dense.

Theorem 1.2 [14] Let A be any subset of a space X . Then $A \in SC(X)$ if and only if $cIA = cIint A$.

The following results can be found in [13].

Proposition 1.3: If a space X is T_1 , then $SO(X) = S_\beta O(X)$.

Proposition 1.4: If a topological space X is locally indiscrete, then every semi-open set is S_β -open set.

Proposition 1.5: A space X is hyper-connected if and only if $S_\beta O(X) = \{X, \emptyset\}$.

Corollary 1.6: For any space X , $S_\beta O(X, \tau) = S_\beta O(X, \tau_\alpha)$

Proposition 1.7: If a topological space (X, τ) is β -regular, then $\tau \subseteq S_\beta O(X)$

Proposition 1.8: 1-Every S_β -open set is S_β -open .

2- An S_β -open set is regular β -open .

3- A regular closed set is S_β -open .

4- Every Regular open set is S_β -closed .

Definition 1.9: A subset A of a space X is said to be S_β -open [18] (resp., S_c [4]) if for each $x \in A \in SO(X)$ there exists a pre-closed (resp., closed) set F such that $x \in F \subseteq A$.

Definition 1.10: A filter base \mathfrak{F} in a space X is s -converges [20] (resp., rc -converges [11]) to a point x if for every semi-open (resp., regular closed) subset U of X containing x there exist $B \in \mathfrak{F}$ such that $B \subseteq cIU$ (resp., $B \subseteq U$).

Definition 1.11: A filter base \mathfrak{F} in a space X is s -accumulates [20] (resp., rc -accumulates [11]) to a point x if for every semi-open (resp., regular closed) subset U of X containing x there exist $B \in \mathfrak{F}$ such that $B \cap cIU \neq \emptyset$ (resp., $B \cap U \neq \emptyset$).

2 S_β -compact and S_β -closed spaces

In this section, we introduce new classes of topological space called S_β -compact and S_β -closed spaces .

Definition 2.1: A filter base \mathfrak{F} is S_β -convergent (resp., S_β - θ -convergent) to a point $x \in X$, if for any S_β -open set V containing x , there exists $F \in \mathfrak{F}$ such that $F \subseteq V$ (resp., $F \subseteq S_\beta cIV$).

Definition 2.2: A filter base \mathfrak{F} is S_β -accumulates (resp., S_β - θ -accumulates) to a point $x \in X$, if $F \cap V \neq \emptyset$ (resp.,

$F \cap S_\beta cIV \neq \emptyset$) for any S_β -open set V containing x and every $F \in \mathfrak{F}$.

It is clear from the definition above, that S_β -converges (resp., S_β -accumulates) of a filter bases in a topological spaces implies S_β - θ -converges (resp., S_β - θ -accumulates) but the converses are not true in general as shown in the following examples

Example 2.3: Let $X = \{a, b, c, d\}$ and $\tau = \{\emptyset, X, \{a, b\}, \{a, b, c\}\}$ we get $S_\beta O(X) = \{\emptyset, X, \{a, b\}, \{a, b, c\}, \{a, b, d\}\}$ and let $\mathfrak{F} = \{X, \{c, d\}, \{\{b, c, d\}\}$, then \mathfrak{F} S_β - θ -converges to a point a , but \mathfrak{F} does not S_β -converges to a because the set $\{a, b\} \in S_\beta O(X)$ contains a , but there exists no $B \in \mathfrak{F}$ such that $B \subseteq \{a, b\}$. Also \mathfrak{F} S_β - θ -accumulates to a point b , but \mathfrak{F} does not S_β -accumulates to b , because the set $\{a, b\} \in S_\beta O(X)$ contains b , but there exist $\{c, d\} \in \mathfrak{F}$ such that $\{a, b\} \cap \{c, d\} = \emptyset$.

Theorem 2.4: Let (X, τ) be a topological space and let \mathfrak{F} be a filter base on X . Then the following statements are equivalent:

- 1- There exists a filter base finer than $\{U_x\}$, where $\{U_x\}$ is the family of S_β -open sets of X containing x .
- 2- There exists a filter base \mathfrak{F}_1 finer than \mathfrak{F} and S_β -converges to x .

Proof: Let \mathfrak{F}_1 be a filter base which is finer than both \mathfrak{F} and $\{U_x\}$. Then \mathfrak{F} S_β -converges to x since it contains $\{U_x\}$.

Conversely, let \mathfrak{F}_1 be the filter base which is finer than \mathfrak{F} and which converges to x . Then \mathfrak{F} must contain $\{U_x\}$ by definition.

Corollary 2.5: If \mathfrak{F} is a maximal filter base in a topological space (X, τ) , then \mathfrak{F} S_β -converges (resp., S_β - θ -converges) to a point $x \in X$ if and only if \mathfrak{F} S_β -accumulates (resp., S_β - θ -accumulates) to a point x .

Proof: Let \mathfrak{F} be a maximal filter base in X and S_β -accumulates to a point $x \in X$, and then by Theorem 2.4, there exists a filter base \mathfrak{F}_1 finer than \mathfrak{F} and S_β -converges to x .

But \mathfrak{F} is maximal filter base. Thus it is S_β -convergent to x .

Theorem 2.6: Let \mathfrak{F} be a filter base in a topological space (X, τ) . If \mathfrak{F} S_β -converges (resp., S_β -accumulates) to a point $x \in X$, then \mathfrak{F} rc -converges (resp., rc -accumulates) at a point $x \in X$.

Proof: Suppose that \mathfrak{F} be a filter base S_β -converges to a

point $x \in X$. Let V be any regular closed set containing x , by Theorem 2.1.15, then V is S_β -open set containing x . since \mathfrak{F} S_β -converges (resp., S_β -accumulates) to a point $x \in X$, There exists $F \in \mathfrak{F}$ such that $F \subseteq V$ (resp., $F \cap V \neq \emptyset$). This shows that \mathfrak{F} rc-converges (resp., rc-accumulates) to a point $x \in X$.

The following example Show that the converse of Theorem 2.6 is not true in general.

Example 2.7: Let $X = \{a, b, c, d\}$ and $\tau = \{\emptyset, X, \{a, c\}\}$, then the family $RC(X) = \{\emptyset, X\}$ and $S_\beta O(X) = \{\emptyset, X, \{a, c\}, \{a, b, c\}, \{a, c, d\}\}$ and $\mathfrak{F} = \{\{b\}, \{a, b, c\}, X\}$. Then \mathfrak{F} is rc-converges (resp., rc-accumulates) to a point $c \in X$, but not S_β -converges (resp., S_β -accumulates) to a point $c \in X$.

Theorem 2.8: Let \mathfrak{F} be a filter base in a topological space (X, τ) . If \mathfrak{F} S_β -converges (resp., S_β -accumulates) to a point $x \in X$, then \mathfrak{F} s-converges (resp., s-accumulates) at a point $x \in X$.

Proof: Suppose that \mathfrak{F} is S_β -converges to a point $x \in X$. Let V be any semi-open set containing x , then by Theorem 1.2 $cIV = cI \text{int } V$, so cIV is regular closed, by Theorem Proposion 1.8 cIV is S_β -open set containing x . Since \mathfrak{F} is S_β -converges (resp., S_β -accumulates) to a point $x \in X$, then there exists $B \in \mathfrak{F}$ such that $B \subseteq cIV$ (resp., $B \cap cIV \neq \emptyset$). This implies that \mathfrak{F} is s-converges (resp., s-accumulates) at a point $x \in X$.

The converse of Theorem 2.8 is not true in general as shown in the following example:

Example 2.9: In Example 2.7 the family $SO(X) = \{\emptyset, X, \{a, b\}, \{a, b, c\}, \{a, b, d\}\} = S_\beta O(X)$, considering the family $\mathfrak{F} = \{\{c\}, \{a, b, c\}, X\}$. Then \mathfrak{F} is filter base s-converges (resp., s-accumulates) to a point $b \in X$, but not S_β -converges (resp., S_β -accumulates) to a point $b \in X$.

Theorem 2.10: Let \mathfrak{F} be a filter base in a topological space (X, τ) and E is any β -closed set containing x . If there exists $F \in \mathfrak{F}$ such that $F \subseteq E$ (resp., $F \subseteq S_\beta cIE$). Then \mathfrak{F} S_β -converges (resp., S_β - θ converges) to a point $x \in X$.

Proof: Let V be any S_β -open set containing x . then V is semi-open and for each $x \in V$, there exist a β -closed set E such that $x \in E \subseteq V$. By hypothesis, there exists $F \in \mathfrak{F}$ such that $F \subseteq E \subseteq V$ (resp., $F \subseteq S_\beta cIE \subseteq S_\beta cIV$). Which implies that

$F \subseteq V$ (resp., $F \subseteq S_\beta cIV$). Hence \mathfrak{F} S_β -converges (resp., S_β - θ converges) to a point $x \in X$.

Theorem 2.11: Let \mathfrak{F} be a filter base in a topological space (X, τ) and E is any β -closed set containing x . If there exists $F \in \mathfrak{F}$ such that $F \cap E \neq \emptyset$ (resp., $F \cap \beta_s cIE \neq \emptyset$), then \mathfrak{F} S_β -accumulates (resp., S_β - θ -accumulates) at a point $x \in X$.

Proof: Similar to proof of Theorem 2.10.

Definition 2.12: A topological space (X, τ) is said to be S_β -compact (resp., S_β -closed) if for every cover $\{V_\alpha : \alpha \in \nabla\}$ of X , by S_β -open sets, there exists a finite subset ∇_0 of ∇ such that $X = \cup\{V_\alpha : \alpha \in \nabla_0\}$ (resp., $X = \cup\{S_\beta cIV_\alpha : \alpha \in \nabla_0\}$)

It is clear that S_β -compact is S_β -closed, but not conversely as shown in the following examples

Example 2.13: Consider an countable space X with Co-countable topology. Since X is T_1 , then by Proposition 1.3 the family of open, semi-open set and S_β -open sets are identical. Hence X is an S_β -closed because every open set in X is dense, not S_β -compact [19] p-194.

Theorem 2.14: If every β -closed cover of a space has a finite subcover, then X is S_β -compact.

Proof: Let $\{V_\alpha : \alpha \in \nabla\}$ be any S_β -open cover of X , and $x \in X$, then for each $x \in V_{\alpha(x)}$ and $\alpha \in \nabla$, there exists a β -closed sets $x \in X$ such that $x \in F_{\alpha(x)} \subseteq V_{\alpha(x)}$, so the family $\{F_{\alpha(x)} : \alpha \in \nabla\}$ is a β -closed cover of X , then by hypothesis, this family has a finite subcover such that $X = \cup\{F_{\alpha(x)} : i = 1, 2, \dots, n\} \subseteq \cup\{V_{\alpha(x)} : i = 1, 2, \dots, n\}$. Therefore $X = \cup\{V_{\alpha(x)} : i = 1, 2, \dots, n\}$. Hence X is S_β -compact.

The following theorem shows the relation between S_β -compact and some other compactness

Theorem 2.15: Every semi-compact space, is S_β -compact spaces.

Proof: Let $\{V_\alpha : \alpha \in \nabla\}$ be any S_β -open cover of X . Then $\{V_\alpha : \alpha \in \nabla\}$ is a semi--open cover of X . Since X is semi-compact, there exists a finite subset ∇_0 of ∇ such that $X = \cup\{V_\alpha : \alpha \in \nabla_0\}$. Hence X is S_β -compact.

The following example shows that the converse of Theorem 2.15 is not true in general.

Example 2.16: Let $X = \mathbb{R}$ with the topology $\tau = \{X, \phi, \{0\}\}$.

Then (X, τ) is not semi-compact, since the space X is hyperconnected, then by Proposition 1.5

$S_\beta O(X) = \{\phi, X\}$. Then (X, τ) is S_β -compact.

Theorem 2.17: If a topological space (X, τ) is T_1 and S_β -compact space, then it is semi-compact.

Proof: Suppose that X is T_1 and S_β -compact space. Let $\{V_\alpha : \alpha \in \nabla\}$ be any semi-open cover of X . Then for every $x \in X$, there exist $\alpha(x) \in \nabla$ such that $x \in V_{\alpha(x)}$. Since X is T_1 by proposition 1.3, the family $\{V_\alpha : \alpha \in \nabla\}$ is S_β -open cover of X . Since X is S_β -compact, there exists a finite subset ∇_0 of ∇ in X such that $X = \bigcup\{V_\alpha : \alpha \in \nabla_0\}$. Hence X is semi-compact.

Theorem 2.18: If a topological space (X, τ) is locally indiscrete. Then S_β -compact space X is semi-compact.

Proof: Follows from Theorem 1.4.

In general S_β -compact spaces and compact spaces are not comparable as shown in the following examples.

Example 2.19: The one-point compactification of any discrete spaces is not S -closed [10] corollary 4, therefore the space is not S_β -compact, but it is compact.

Example 2.20: Let $X = (0, 1)$ with the topology $\tau = \{X, \phi, G = (0, 1 - \frac{1}{n}), n = 2, 3, \dots\}$ then (X, τ) is not compact [19], p. 76, but it is S_β -compact since the only S_β -open subset of (X, τ) are X and ϕ .

Theorem 2.21: Let (X, τ) be a topological space. If X is β -regular and S_β -compact, then X is compact.

Proof: Let $\{V_\alpha : \alpha \in \nabla\}$ be any open cover of X . Since X is β -regular By Proposition 1.7, $\{V_\alpha : \alpha \in \nabla\}$ forms an S_β -open cover of X . Since X is S_β -compact, there exists a finite subset ∇_0 of ∇ such that $S_\beta \text{int} F_{\alpha_0} \cap S_\beta \text{cl}(X \setminus F_{\alpha_0}) = \phi$, hence X is compact.

Theorem 2.22: If X is an s -regular S_β -closed T_1 -space, then X is compact.

Proof: Let $\{V_\alpha : \alpha \in \nabla\}$ be any open cover of an s -regular and S_β -closed T_1 space X , then for each $\alpha \in \nabla$ and for each $x \in X$ there exists $\alpha(x) \in \nabla$ such that $x \in V_{\alpha(x)}$. Since X is s -

regular T_1 space there exist $G_x \in SO(X)$ and $x \in G_x \subset \text{cl}G_x \subset V_{\alpha(x)}$. Then by Proposition 1.3, the family $\{G_x : x \in X\}$ is an S_β -open cover of X . Since X is S_β -closed space, then there exists a subfamily $\{G_{x_i} : i = 1, 2, \dots, n\}$ such that $X = \bigcup_{i=1}^n \text{cl}G_{x_i} \subseteq \bigcup_{i=1}^n V_{\alpha(x_i)}$. Thus X is compact.

The following example shows that the condition of s -regularity in Theorem 2.22 can not be dropped

Example 2.23: In Example 2.20 (X, τ) is neither s -regular nor compact but it is S_β -closed because the only non-empty S_β -open set in (X, τ) is X itself.

Theorem 2.24: Let X be an almost-regular space if X is S_β -compact, then it is nearly compact.

Proof: Let $\{V_\alpha : \alpha \in \nabla\}$ be any regular open cover of X , Since X is almost-regular space then for each $x \in X$ and each regular open set $V_{\alpha(x)}$, there exist an open set G_x such that

$x \in G_x \subset \text{cl}G_x \subset V_{\alpha(x)}$. But $\text{cl}G_x$ is regular closed for each $x \in X$. Therefore $X = \bigcup_{x \in X} \text{cl}G_x = \bigcup_{\alpha(x) \in \nabla} V_{\alpha(x)}$ This implies that the

family $\{\text{cl}G_x : x \in X\}$ is an S_β -open cover of X . Since X is S_β -compact, then there exists a subfamily

$\{\text{cl}G_{x_i} : i = 1, 2, \dots, n\}$ such that $X = \bigcup_{i=1}^n \text{cl}G_{x_i} \subseteq \bigcup_{i=1}^n V_{\alpha(x_i)}$. Thus X

is nearly compact.

Theorem 2.25: Let X be semi-regular s -closed space, then it is S_β -compact.

Proof: Let $\{V_\alpha : \alpha \in \nabla\}$ be any S_β -open cover of X , Then V_α is semi-open for each $\alpha \in \nabla$, since X is semi-regular for each $x \in X$ and $V_{\alpha(x)}$, there exists a semi-open set G_x such that

$x \in G_x \subseteq \text{cl}G_x \subseteq V_{\alpha(x)}$. Then the family $\{G_x : x \in X\}$ is semi-open cover of X . Since X is s -closed space, then there exists a subfamily $\{G_{x_i} : i = 1, 2, \dots, n\}$ such that $X = \bigcup_{i=1}^n \text{cl}G_{x_i} \subseteq \bigcup_{i=1}^n V_{\alpha(x_i)}$.

Thus X is S_β -compact.

Theorem 2.26: For any topological space (X, τ) . The following statements are equivalent:

- 1- (X, τ) is S_β -compact spaces (resp., S_β -closed)
- 2- For any S_β -open cover $\{V_\alpha : \alpha \in \nabla\}$ of X , there exists a finite subset ∇_0 of ∇ such that $X = \bigcup\{V_\alpha : \alpha \in \nabla_0\}$ (resp.,

$$X = \bigcup \{S_\beta cIV_\alpha : \alpha \in \nabla_0\}$$

3- Every maximal filter base \mathfrak{F} in X S_β -converges (resp. S_β - θ -converges) to some point $x \in X$.

4- Every filter base \mathfrak{F} in X S_β -accumulates (resp., S_β - θ -accumulates) to some point $x \in X$.

5- For every family $\{F_\alpha : \alpha \in \nabla\}$ of S_β -closed subsets of X such that $\bigcap \{F_\alpha : \alpha \in \nabla\} = \emptyset$ there exists finite subset ∇_0 of ∇ such that $\bigcap \{F_\alpha : \alpha \in \nabla_0\} = \emptyset$ (resp., $\bigcap \{S_\beta \text{ int } F_\alpha : \alpha \in \nabla_0\} = \emptyset$)

proof: 1 \rightarrow 2. Straightforward.

2 \rightarrow 3 Suppose that for every S_β -open cover $\{V_\alpha : \alpha \in \nabla\}$ of X , there exist a finite subset ∇_0 of ∇ such that

$$X = \bigcup \{V_\alpha : \alpha \in \nabla_0\} \text{ (resp., } X = \bigcup \{S_\beta cIV_\alpha : \alpha \in \nabla_0\}) \text{ and let}$$

$\mathfrak{F} = \{F_\alpha : \alpha \in \nabla\}$ be a maximal filter base. Suppose that \mathfrak{F} does not S_β -converges (resp., S_β - θ -converges) to any point of X .

Since \mathfrak{F} is maximal, by Corollary 5.1.4 \mathfrak{F} does not S_β -accumulates (resp., S_β - θ -accumulates) to any point of X .

This implies that for every $x \in X$ there exists S_β -open set

$$V_x \text{ and } F_{\alpha(x)} \in \mathfrak{F} \text{ such that } F_{\alpha(x)} \cap V_x = \emptyset \text{ (resp.,}$$

$$F_{\alpha(x)} \cap S_\beta cIV_x = \emptyset). \text{ The family } \{V_x : x \in X\} \text{ is an } S_\beta \text{-open cover of}$$

X and by hypothesis, there exists a finite number of points

$$x_1, x_2, \dots, x_n \text{ of } X \text{ such that } X = \bigcup \{V_{(x_i)} : i = 1, 2, \dots, n\} \text{ (resp.,}$$

$$X = \bigcup \{S_\beta cIV_{(x_i)} : i = 1, 2, \dots, n\}). \text{ Since } \mathfrak{F} \text{ is a filter base on } X, \text{ there}$$

exists a $F_0 \in \mathfrak{F}$ such that $F_0 \subseteq \bigcap \{F_{\alpha(x_i)} : i = 1, 2, \dots, n\}$. Hence

$$F_0 \cap V_{\alpha(x_i)} = \emptyset \text{ (resp., } F_0 \cap S_\beta cIV_{\alpha(x_i)} = \emptyset) \text{ for } i = 1, 2, \dots, n \text{ which}$$

implies that $F_0 \cap (\bigcup \{V_{(x_i)} : i = 1, 2, \dots, n\}) = \emptyset$ (resp.,

$$F_0 \cap (\bigcup \{S_\beta cIV_{(x_i)} : i = 1, 2, \dots, n\}) = F_0 \cap X = \emptyset$$

Therefore, we obtain $F_0 = \emptyset$. Which is contradict the fact that $\mathfrak{F} \neq \emptyset$, thus

\mathfrak{F} is S_β -converges to some point $x \in X$

3 \rightarrow 4. Let \mathfrak{F} be any filter base on X . Then, there exists a maximal filter base \mathfrak{F}_0 such that $\mathfrak{F} \subseteq \mathfrak{F}_0$. By hypothesis \mathfrak{F}_0 is S_β -converges (resp., S_β - θ -converges) to some point $x \in X$.

For every $F \in \mathfrak{F}$ and every S_β -open set V containing x , there exists $F_0 \in \mathfrak{F}_0$ such that $F_0 \subseteq V$ (resp., $F_0 \subseteq S_\beta cIV$), hence $\emptyset \neq F_0 \cap F \subseteq V \cap F$ (resp., $S_\beta cIV \cap F$). This shows that \mathfrak{F} S_β -accumulates at x (resp., S_β - θ -accumulates).

4 \rightarrow 5. Let $\{F_\alpha : \alpha \in \nabla\}$ be a family of S_β -closed subsets of X such that $\bigcap \{F_\alpha : \alpha \in \nabla\} = \emptyset$. If possible suppose that every finite subfamily $\bigcap \{F_{\alpha_i} : i = 1, 2, \dots, n\} \neq \emptyset$. Therefore $\mathfrak{F} = \mathcal{A} \subseteq \mathcal{Y} \subseteq X$ form

a filter base on X . By hypothesis, \mathfrak{F} S_β -accumulates (resp., S_β - θ -accumulates) to some point $x \in X$. This implies that for every S_β -open set V containing x , $F_\alpha \cap V \neq \emptyset$ (resp., $F_\alpha \cap S_\beta cIV \neq \emptyset$), for every $F_\alpha \in \mathfrak{F}$ and every $\alpha \in \nabla$. Since $x \notin \bigcap F_\alpha$, there exist $\alpha_0 \in \nabla$ such that $x \notin F_{\alpha_0}$. Hence $X \setminus F_{\alpha_0}$ is S_β -open set containing x and $F_{\alpha_0} \cap X \setminus F_{\alpha_0} = \emptyset$ (resp., $S_\beta \text{ int } F_{\alpha_0} \cap S_\beta cIV(X \setminus F_{\alpha_0}) = \emptyset$). Which contracting the fact that \mathfrak{F} S_β -accumulates to x (resp., S_β - θ -accumulates). So the assertion in (5) is true.

5 \rightarrow 1. Let $\{V_\alpha : \alpha \in \nabla\}$ be S_β -open cover of X .

Then $\{X \setminus V_\alpha : \alpha \in \nabla\}$ is a family of S_β -closed subsets of X such that $\bigcap \{X \setminus V_\alpha : \alpha \in \nabla\} = \emptyset$. By hypothesis, there exists a finite subset ∇_0 of ∇ such that $\bigcap \{X \setminus V_\alpha : \alpha \in \nabla_0\} = \emptyset$ (resp., $\bigcap \{S_\beta \text{ int}(X \setminus V_\alpha) : \alpha \in \nabla_0\} = \emptyset$). Hence $X = \bigcup \{V_\alpha : \alpha \in \nabla_0\}$ (resp., $X = \bigcup \{S_\beta cIV_\alpha : \alpha \in \nabla_0\}$). This shows that X is S_β -compact (resp., S_β -closed).

Theorem 2.27: If a topological space (X, τ) is S_β -closed and T_1 -space then it is nearly compact.

Proof: Let $\{V_\alpha : \alpha \in \nabla\}$ be any regular open cover of X . Then $\{V_\alpha : \alpha \in \nabla\}$ is a S_β -open cover of X . Since X is S_β -closed, there exists a finite subset ∇_0 of ∇ such that $X = \bigcup \{V_\alpha : \alpha \in \nabla_0\}$. Hence X is nearly compact.

3 Characterization of S_β -compact spaces

Definition 3.1: A point x in X is said to be S_β -complete

accumulation point of a subset A of X if

$\text{Card}(A \cap U) = \text{Card}(A)$ for each $U \in S_\beta(X, x)$. Where $\text{Card}(A)$ denotes the cardinality of A .

Definition 3.2: In a topological space X , a point x is said to be an S_β -adherent point of a filter \mathfrak{F} on X if it lies in the S_β -closure of all sets of \mathfrak{F} .

Theorem 3.3: A space X is S_β -compact spaces if and only if each infinite subset of X has S_β -complete accumulation point.

Proof: Let the space X be S_β -compact and S an infinite subset of X . Let K be the set of points x in X which are not S_β -complete accumulation points of S . Now it is obvious that for each point x in K , we are able to find $U_{(x)} \in S_\beta O(X, x)$ such that $\text{Card}(S \cap U_{(x)}) \neq \text{Card}(S)$. If K is the Whole space, then

$\mathfrak{R} = \{U_{(x)} : x \in X\}$ is S_β -cover of X . By hypothesis X is S_β -compact, so there exists a finite subcover $\Psi = \{U_{(x_i)}; i = 1, 2, \dots, n\}$ such that $S \subseteq \cup\{U_{(x_i)} \cap S; i = 1, 2, \dots, n\}$, then $Card(S) = \max\{Card(U_{(x_i)} \cap S); i = 1, 2, \dots, n\}$, which does not agree with what we assumed. This implies that S has an S_β -complete accumulation. Now assume that X is not S_β -compact and that every infinite subset S of X has an S_β -complete accumulation point in X . it follows that there exists an cover Θ with no finite subcover. Set $\delta = \min\{Card(\Xi); \Xi \subset \Theta, \text{wher } \Xi \text{ is an } S_\beta\text{-cover of } X\}$. Fix $\Psi \subseteq \Theta$, for which $Card(\Psi) = \delta$ and $\cup\{U : U \in \Psi\} = X$. Let N denote the set of natural numbers, then by hypothesis $\delta \geq Card(N)$ [By well-ordering of Ψ]. By some minimal well-ordering " \sim ", suppose that U is any member of Ψ . By minimal well-ordering " \sim ", we have $Card(\{V : V \in \Psi, V \sim U\}) < Card(\{V : V \in \Psi\})$. Since Ψ can not have any subcover with cardinality less than δ , then for each $U \in \Psi$ we have $X \neq \cup\{V : V \in \Psi, V \sim U\}$. For each $U \in \Psi$ choose a point $x(U) \in X - \cup\{V \cup \{x(V)\}; V \in \Psi, V \sim U\}$. We are always able to do this, if not, one can choose a cover of smaller cardinality from Ψ . If $H = \{x(U); U \in \Psi\}$, then to finish the proof we will show that H has no S_β -complete accumulation point in X . Suppose that z is a point of the space X . Since Ψ is S_β -cover of X , then z is a point of some set W in Ψ . By the fact that $U \sim V$ we have $x(U) \in W$. It follows that $T = \{U : U \in \Psi \text{ and } x(U) \in W\} \subset \{V : V \in \Psi, V \sim W\}$. But $Card(T) < \delta$. Therefore $Card(H \cap W) < \delta$. But $Card(H) = \delta \geq Card(N)$, since for two distinct points U and W in Ψ , we have $x(U) \neq x(W)$, this means that H has no S_β -complete accumulation point in X which contradicts our assumptions. Therefore X is S_β -compact.

Theorem 3.4: For a topological space the following are equivalent:

- i- X is S_β -compact.
- ii- Every net in X with well-ordered directed set as its domain accumulates to some point of X .

Proof: (i) \Rightarrow (ii): Suppose that X is S_β -compact and

$\xi = \{x_\alpha : \alpha \in \nabla\}$ a net with a well-ordered set ∇ as domain. .

Assume that ξ has no S_β -adherent point in X . Then for each point x in X there exists $V_{(x)} \in S_\beta O(X, x)$ and an

$\alpha(x) \in \nabla$ such that $V_{(x)} \cap \{x_\alpha : \alpha \geq \alpha(x)\} = \emptyset$. This implies that $\{x_\alpha : \alpha \geq \alpha(x)\}$ is a subset of $X - V_{(x)}$. Then the collection $\omega = \{V_{(x)} : x \in X\}$ is S_β -cover of X . By hypothesis of theorem, X is S_β -compact and so ω has a finite subfamily $\{V_{(x_i)} : i = 1, 2, \dots, n\}$ such that $X = \cup\{V_{(x_i)} : i = 1, 2, \dots, n\}$. Suppose that the corresponding elements of ∇ be $\{\alpha(x_i)\}$ where $i = 1, 2, \dots, n$, since ∇ is well-ordered and $\{\alpha(x_i)\}$ where $i = 1, 2, \dots, n$ is finite. The largest elements of $\{\alpha(x_i)\}$ exists. Suppose it is $\alpha(x_i)$. Then for $\gamma \geq \alpha(x_i)$. We have $\{x_\delta : \delta \geq \gamma\} \subset \bigcap_{i=1}^n (X - V_{(x_i)}) = X - \bigcup_{i=1}^n V_{(x_i)} = \emptyset$. Which is impossible.

This shows that ξ has at least one S_β -adherent point in X .

(ii) \Rightarrow (i): Now it is enough to prove that each infinite subset has an S_β -complete accumulation point by utilizing above theorem. Suppose that $S \subseteq X$ is an infinite subset of X .

According to Zorn's Lemma, the infinite set S can be well-ordered. This means that we can assume S to be a net with a domain which is a well ordered index set. It follows that S has S_β -adherent point z . Therefore is an S_β -complete accumulation point of S this shows that X is S_β -compact.

Theorem 3.5 : A space X is S_β -compact if and only if each family of S_β -closed subsets of X with the finite intersection property has a non-empty intersection .

Proof: Given a collection ω of subsets of X . let $\nu = \{X - w : w \in \omega\}$ be the collection of their complements . Then the following statements hold.

- i- ω is the collection of S_β -open sets if and only if ν is a collection of S_β -closed sets.
- ii- the collection ω covers of X if and only if the intersection $\bigcap_{v \in \nu} v$ of all the elements of ν is non empty
- iii- The finite sub collection $\{w_1, \dots, w_n\}$ of ω covers X if and only if the intersection of the corresponding elements $v_i = X - w_i$ of ν is empty.

The statement (i) is trivial . While the statement (ii) and (iii) follows from De-Morgan Law $X - \bigcup_{\alpha \in J} v_\alpha = \bigcap_{\alpha \in J} (X - v_\alpha)$. The

proof of theorem now proceeds in two steps. Taking the contra positive of the theorem and the complement .

The statement X is S_β -compact is equivalent to: Given any collection ω of S_β -open subsets of X , if ω covers X , then

some finite sub collection of ω covers X . This statement is equivalent to its contra positive, Which is the following.

Given any collection ω of S_β -open sets, if no finite sub collection of ω covers X , then ω does not cover X . Letting \mathcal{V} be as earlier, the collection $\{X - w : w \in \omega\}$, and applying (i) to (iii), we see that this statement is in turn equivalent to the following.

Given any collection \mathcal{V} of S_β -closed sets, if every finite intersection of elements of \mathcal{V} is non empty. This is just the condition of our theorem.

Theorem3.6: A space X is S_β -compact if and only if each filter base in X has at least one S_β -adherent point.

Proof: Suppose that X is S_β -compact and $\mathfrak{F} = \{F_\alpha : \alpha \in \nabla\}$ is a filter base in it. Since all finite intersections of F_α 's are nonempty. It follows that all finite intersections of $S_\beta cl(F_\alpha)$'s are also nonempty. Now it follows from Theorem [F.I.P] that $\bigcap_{\alpha \in \nabla} S_\beta cl F_\alpha$ is nonempty. This means that \mathfrak{F} has at least one S_β -adherent point. Now suppose that \mathfrak{F} is any family of S_β -closed sets. Let each finite intersection be nonempty the set F_α with their finite intersection establish the filter base \mathfrak{F} . Therefore \mathfrak{F} S_β -accumulates to some point z in X . It follows that $z \in \bigcap_{\alpha \in \nabla} F_\alpha$. Now we have by Theorem 3.5, that X is S_β -compact.

Theorem3.7: A space X is S_β -compact if and only if each filter base on X , with at most one S_β -adherent point, is S_β -convergent.

Proof: Suppose that X is S_β -compact, x is a point of X , and \mathfrak{F} is a filter base on X . The S_β -adherent of \mathfrak{F} is a subset of $\{x\}$. Then the S_β -adherent of \mathfrak{F} is equal to $\{x\}$, by Theorem 3.7. Assume that there exists a $V \in S_\beta O(X, x)$ such that for all $F \in \mathfrak{F}$, $F \cap (X - V)$ is nonempty. Then $\Psi = \{F - V : F \in \mathfrak{F}\}$ is a filter base on X . It follows that the S_β -adherence of Ψ is nonempty. However

$$\bigcap_{F \in \Psi} S_\beta cl(F - V) \subseteq \left(\bigcap_{F \in \mathfrak{F}} S_\beta cl F \right) \cap (X - V) = \{x\} \cap (X - V) = \emptyset.$$

But this is a contradiction. Hence, for each $V \in S_\beta O(X, x)$ there exist

$F \in \mathfrak{F}$ with $F \subseteq V$. This shows that \mathfrak{F} S_β -converges to x .

To prove the converse, It suffices to show that each filter base in X has at least one S_β -accumulation point. Assume that \mathfrak{F}

is a filter base on X with no S_β -adherent point. By hypothesis

\mathfrak{F} S_β -converges to some point z in X . Suppose F_α is an arbitrary element of \mathfrak{F} . Then for each $V \in S_\beta O(X, z)$, there exists an $F_\beta \in \mathfrak{F}$ such that $F_\beta \subseteq V$. Since \mathfrak{F} is a filter base there exists a γ such that $F_\gamma \subseteq F_\alpha \cap F_\beta \subseteq F_\alpha \cap V$ where F_γ is a nonempty. This means that $F_\alpha \cap V$ is nonempty for every $V \in S_\beta O(X, z)$ and correspondingly for each α , z is a point of $S_\beta cl F_\alpha$. It follows that $z \in \bigcap_{\alpha} S_\beta cl F_\alpha$. Therefore, z is

S_β -adherent point of \mathfrak{F} . Which is contradiction. This shows that X is S_β compact.

REFERENCES

- [1]. M.E.Abd El-Monsef, S.N. El-Deeb and R.A Mahmoud., β -open sets and β -continuous mappings, Bull. Fac. Sci. Assiut Univ., 12(1983), 77-90.
- [2]. N.K. Ahmed, On some types of separation axioms, M. Sc. Thesis, College of Science, Salahaddin Univ. 1990.
- [3]. Z. A. Ameen, A new class of semi-open sets in topological spaces, M. Sc. Thesis, College of Science, Dohuk Univ. 2009.
- [4]. B.A. Asaad, Utilization of some types of pre-open sets in topological space, M. Sc. Thesis, College of Science, Dohuk Univ. 2007.
- [5]. D.E.Cameron, Properties of S-closed spaces, Proc. Amer. Math. Soc. 72 (1978), 581-586.
- [6]. S.G. Crossely and S.K Hildebrand., Semi closure, Texas. J. Sci., 22 (1971), 99-112.
- [7]. G.Di maio and T.Noiri, On S-closed spaces, Indian J. Pure Appl. Math., 18(3) (1987), 226-233.
- [8]. K.Dlaska and M.Ganster, S-sets and Co-S-closed topologies, Indian J. Pure Appl. Math., 23(10)(1992), 731-737.
- [9]. J.Dontchev, Survey on pr-eopen sets, The proceedings of Yatsushiro topological conference, (1998), 1-18
- [10]. J. Dontchev, M. Ganster and T. Noiri, On P-closed spaces, Internet. J.Math. and Math Sci., 24(3)(2000), 203-212.
- [11]. R.A. Hermann, rc-convergence, Proc. Amer. Math.Soc., Vol.75, No.2, (1979) 311-317.
- [12]. J.E. Joseph and M.H. Kwach, On S-closed spaces, Proc. Amer. Math Soc. 80(2)(1980), 341-348.

- [13] A.B Khalaf and N.K. Ahmed , S_β -open set and S_β - continuous function, submitted.
- [14] N. Levine, Semi open sets and semi-continuity in topological spaces, Amer.Math. Monthly, 70 (1963), 36 –41.
- [15] A.S. Mashhour, M.E Abd El-monsef, and S.N. El.Deeb, On pre- continuous and weak pre-continuous mappings, Proc. Math and phys. Soc. Egypt 53 (1982), 47-53.
- [16] O.Njåstad, On some classes of nearly open sets, Pacific J. Math., 15(3) (1965),961- 970.
- [17] T. Noiri , On semi continuous mapping, Accad. Naz. Lincei. Rend. CL. Sci. Fis. Mat. Natur., Vol54,No. 8(1973)210-214.
- [18] A.H. Shareef , S_p -open sets, S_p -continuity and S_p - compactness in topological spaces, M. Sc. Thesis ,College of Science ., Sulaimani Univ.2007.
- [19]. L.A Steen and J.A. Seebach , Counterexamples in topology, Holt, Rinehart and Winston, Inc., U.S.A., 1970.
- [20] T. Thompson , S-closed spaces , , Proc. Amer. Math.Soc.,Vol.60, No.1, (1976)335-338.
- [21] N.V. Velicko , H-closed topological spaces, Amer. Math. Soc. Transl. 78(2) (1992), 103-118.
- [22] R.H.Yunis , Regular β -open sets, Zanco J. of pure and applied Science, 16(3) (2004) 79-83

Barriers and benefits of implementing a power quality program: Case Study Libyan distribution networks

Saad S. Sultan, Mohamed K. Darwish

Abstract— The purpose of this paper is to explore the obstructions faced by Libyan distribution networks in implementing a power quality program (PQP). It is also to state the benefits, which would accrue by implementing a PQP, which would make a major impact on Libyan distribution networks (LDNs), and which could be applied and adapted internationally. In order to achieve these objectives, an extensive literature review was conducted to understand the barriers and benefits of implementing a PQP, followed by a power quality survey questionnaire and interviews. Data were collected from LDNs, both from departments and individual staff members. Both SPSS 15.1 and Nvivo 9 were used in performing the analysis. The results revealed that no power quality program exists. Out of 16 barriers, 12 were statistically significant different since the P value $<.05$, which indicated that Libya distribution systems have already surmounted a few of the barriers to implementing a PQP effectively. The overall benefits of PQP implementation, which would have a positive impact on LDNs, are 11 benefits. Improving power quality disturbances (PQDs) and achieving the objectives of the implementation of PQP are, influenced by the distribution networks in tackle the obstacles, which remain. The findings of a LDN survey were compared with other studies and suggestions were made for the future improvement.

Index Terms— Barriers, Benefits, Libyan Distribution Networks, PQP Framework

◆

1 INTRODUCTION

Power quality programs have become some of the most recent services among distribution companies, both for private and state suppliers [1], [2]. For any distribution system to satisfy its consumers, the utility must keep improving power quality in a way, which that accommodates the increased demand for electricity [3], [4]. This requires the PQP to be implemented to start tackling the difficulties faced the distribution utilities in sustaining a high standard of power quality. A PQP can help in reducing the huge number of complaints from end users, and the costs represented in the damage to their equipment [5]. It can also have a positive impact on the electrical distribution companies, improving their service and saving some of the significant resources spent. Therefore, the distribution companies need to implement a PQ investigation program despite all the previous facts indicating an increase in PQDs, in the last two decades in particular [6].

Without establishing a clear vision of these barriers, such as the lack of the following: a clear strategy, customer awareness, accommodating economic growth, equipments standards, network design, resources, staff awareness, top management responsibility and power quality standards, together with an excessive increase of electronic equipments, then any efforts to improve power quality will be wasted in both time and resources. Accordingly, several less developed countries have compelled their utilities to implement PQPs, which are offered as a mandatory service, in response to the high increase in customer complaints [7], [8]. This is caused by the increase in sophisticated industrial and commercial equipments, while customers do not demand PQ standards to run it [9].

LDNs are among those systems facing poor power quality in under-developed countries. Unfortunately, statistical data show that in the last two decades, LDNs have not implemented power quality program [10], [11]. This is mainly because there is no power quality department established yet, to influence the measurement of PQDs. This absence of a power quality department is due to lack of awareness on the part of top management regarding the importance of power quality. As a result, LDNs have faced very significant difficulties in implementing PQP. In addition, lack of power quality awareness has led LDNs to face twelve significant difficulties through not implementing PQP [12], [13].

2 POWER QUALITY PROGRAMS (PQPs)

PQPs are particularly successful in developed countries rather than developing countries, due to the rapid adoption of sophisticated technology, as well as the higher level of PQ awareness among most of the end users, who recognize its importance. Furthermore, power suppliers in developed regions are trying to establish a high level of PQ standards in a short time, due to pressure from large industrial customers, as the use of sophisticated equipment increases [6].

In contrast, utilities in less developed countries are being pushed by the introduction of new technology from developed countries to improve and address their PQ issues. In response, some distribution companies have contracted a third party to solve PQ issues for their end users satisfaction; this is due to the inability of their engineers and technicians, who lack the skills and experience to solve these problems [2]. Therefore, government-controlled utilities are detached from the situation

with regard to PQ issues. The failure to implement PQPs by some distribution utilities in developing countries have resulted in their supplying free power to their customers.

The distribution utilities in less developed countries are not worried about the quality of the power they provide to their clients. They believe that PQ has matured to the point, where it will not be of any importance in the future; moreover, their customers want only to be supplied with electricity, and are

not concerned over quality [14]. Therefore, managers from distribution companies have concluded that some international electricity companies view implementing PQP as a business, rather than concerning themselves with issues of power distribution systems [2]. As a result, table 1 states the cost of industries and end users suffer losses due to poor power quality and the failure of implementing PQP.

TABLE 1
COST OF INDUSTRIES AND END USERS LOSSES DUE TO POOR POWER QUALITY

Country	PQ Disturbances	Total Cost annually	Aut hors	Methodology
Brazil	Harmonics Voltage sag, Power interruption	\$ 1.2 m	[15]	Case study
Italy	Voltage sag	\$ 235m	[16]	Case study
New York	Voltage sags	\$ 1 m	[17]	Case study
China	Power interruption	\$ 1.5 m	[18]	Presentation
Germany	Voltage sag, Power interruption	€32 bn	[19]	Panel Discussion
8 Developed Countries Austria, France, Italy, Poland, Portugal, Slovenia, Spain and UK	Voltage dips, Short interruptions, Long interruptions, Harmonics, Transients, and surges	€150 bn	[20]	Interviews and Questionnaire
Massachusetts USA	Voltage fluctuations, Voltage sag	\$ 1.4 bn	[21]	Case study, Interviews
USA	Voltage dips, Short interruptions	\$ 119 to \$ 188 bn	[22]	Survey
Taiwan	Voltage dip	€ 1.7 m	[23]	Case study
Singapore	Short interruption, Voltage dip	€ 3 m	[24]	Survey
Sweden	Voltage dip	€ 2.4 m	[25]	Survey
California industries sectors	Harmonic, Voltage sag	\$ 18.8 bn	[26]	Survey
France	Harmonic, Power interruption, Voltage dips	€1 m	[27]	Survey
UK	Transients, Interruption	£ 200 m	[28]	Insurance Compensation

To classify the barriers facing the implementation of a PQP and also the expected benefits from implementing such a program, a literature review has been carried out and is summarized in two sections, namely PQP barriers and PQP benefits:

3 BARRIERS TO IMPLEMENTING A PQP

Since 1980, PQ issues have been causing real and significant disturbances to the distribution systems and end users worldwide, becoming a global concern [29], [30], [31], [32], [33], [34], [35], [36]. Hence, the lack of awareness of PQ could result in utilities still suffering from PQ problems caused by end users' sensitive equipment for industrial, agriculture, residential and commercial [3]. Therefore, providing sufficient introduction, definitions and explanations for the most widespread PQ terms, will help in identifying the more common PQ disturbances that occur. Moreover, those producing or using the power, in particular in less developed countries, should understand what PQ means.

The reason is that as long as the concept of PQ is misunderstood by both the staff of the electrical distribution

company and the end users, then the severity of PQ issues will increase every day, because the demand for power will increase and even double [1].

Several authors and researchers have determined different aspects of barriers according to their experience and their studies on the implementation of PQP.

A study in the UK revealed eight major categories of PQP barriers: lack of staff awareness regarding PQ issues; lack of enough resources; lack of PQ training courses; lack of top management committed to implementing good PQP; lack of long-term strategy for successful implementation; lack of end users' awareness; lack of PQ standards and lack of regular maintenance [37].

A study conducted by Ghatol and Kushare found two aspects of PQP barriers in less developed countries; lack of network designing; and lack of end users' awareness regarding power quality [38]. A survey in the USA, conducted for the North American Delivery Systems found two barriers to PQP implementation; lack of customer cooperation i.e. illegal connection made by end users; and lack of top management responsibility to face customer complaints [39]. A study in a Massachusetts distribution system found three

barriers to PQP implementation; lack of PQ standards; lack of cooperation by end users; and lack of management commitment regarding end users' complaints [21].

A study by EPRI in the USA pointed out nine components of PQP implementation barriers; lack of top management commitment, support and encouragement; lack of skills, knowledge and experience among engineers' and technicians; lack of proper teams to analyse PQ disturbances; lack of training courses; and lack of a PQ database [40]. Another study in the USA, Asia, Africa, Australia, South America and Europe revealed a lack of power quality awareness among end users; and lack of PQ training courses [2]. A further study in the USA revealed two barriers believed to hinder the successful implementation of PQP; lack of a utilities distribution structure; and lack of suitable management structure and operation [34].

A study in Malaysia found that five barriers to implementing a PQP were a ; lack of education programs; lack of PQ awareness and guidelines; lack of training courses and support; lack of continuing research and development; and lack of financial incentives to encourage the staff to resolve PQ issues [8].

A survey conducted in 8 developed European countries, namely; Austria, France, Italy, Poland, Portugal, Slovenia, Spain and the UK; found that a lack of end users' awareness; lack of employee awareness and skills; lack of management commitment; and lack of PQ measurements and maintenance are the main barriers to PQP implementation. These factors have led to huge economic losses in Europe, exceeding €150bn annually [20]. Another survey in Europe found that the main difficulties encountered during the implementation of PQP are lack of PQ awareness among top management, engineers and end users; lack of network designing, due to increased power demand; lack of PQ standards; lack of PQ measurement [27].

A study in Canada revealed that three main factors impede the wider spread of PQPs; lack of PQ consultants; lack of PQ standards; and lack of PQ awareness on the part of end users [31].

A study in the Netherlands found five significant difficulties in implementing PQP, namely; lack of a

distribution networks infrastructure; failure to handle end users' complaints so as to identify the underlying problems; lack of PQ contracts between suppliers and end users; increasing sensitive electronic equipments; lack of PQ training courses to raise the education and awareness levels of engineers to understand consumers' complaints better [41].

Another study in Germany found twelve barriers to PQP implementation; lack of distribution network designing, structure and size; lack of data on end users' load characteristics and structure; inadequate background and experience among employees regarding PQ; lack of PQ standards; lack of PQ measurement; lack of management planning and strategy [39].

A study in India found two major barriers to PQP implementation; lack of PQ measurement; lack of PQ awareness and skills among employees [42]. A second study in India found four significant categories of PQP barriers; lack of planning and designing the distribution network; lack of proper PQ teams; lack of PQ monitoring and databases to analyze customer complaints; and lack of PQ standards [43]. In Pakistan, a study found that lack of understanding PQ disturbances is a major obstacle to the implementation of a PQP to be achieved [44].

A study conducted by Moncrief, Dougherty, Richardson, and Craven found five main barriers to PQP implementation; lack of end users' awareness; lack of PQ equipment standards; lack of PQ awareness among employees; lack of PQ monitoring and databases regarding end users' complaints as a form of assistance to the utilities; lack of PQ measurements [45]. A study in Latin America found three barriers encountered during the implementation of PQP; lack of PQ monitoring and datasets; lack of PQ standards; lack of PQ employee' awareness and experience [46].

A study in Brazil found seven factors as the main barriers to PQP implementation; lack of distribution networks infrastructure; lack of studies and research; lack of distribution network design; lack of management planning; lack of technician and engineer skills and experience; lack of end users' awareness; lack of a clear strategy [47]. Table 2 shows the different and similar barriers by the above researchers.

TABLE 2

THE DIFFERENT AND SIMILAR OF POWER QUALITY PROGRAM DISCERNED BY THE ABOVE RESEARCHERS

Barriers	Country
lack of staff awareness, skills and experience	USA, European, India, Latin America, Brazil, Germany, Pakistan, Austria, France, Italy, Poland, Portugal, Slovenia, Spain and UK,
lack of enough resources	USA, UK
lack of top management commitment	USA, Massachusetts, Austria, France, Italy, Poland, Portugal, Slovenia, Spain and UK,
lack of long-term strategy and planning	USA, Brazil, Germany, UK
lack of end users awareness	USA, Asia, Africa, Australia, South America and Europe, Canada, Brazil, Austria, France, Italy, Poland, Portugal, Slovenia, Spain and UK,
lack of network designing	USA, European, India, Brazil, Germany
lack of training courses, and support	Malaysia, USA, Asia, Africa, Australia, South America and Europe, Netherlands, UK
lack of conducting research and studies	Malaysia, Brazil
lack of financial incentives	Malaysia

lack of customer cooperation	USA, Massachusetts
lack of top management responsibility	USA, Netherlands
lack of PQ standards	Massachusetts , European, Canada, India, Latin America, Germany, UK
lack of PQ measurement	India, USA, Germany, Austria, France, Italy, Poland, Portugal, Slovenia, Spain, UK,
lack of PQ consultants	Canada, India, USA
lack of DNs infrastructure	Netherlands, Brazil
lack of PQ monitoring and database	India, USA, Latin America
lack of regular maintenance	UK, Austria, France, Italy, Poland, Portugal, Slovenia, Spain

4 BENEFITS OF IMPLEMENTING PQP

Power quality program, if effectively implemented, will lead to substantial benefits. Tackling the barriers, to a high level of PQ and high level of end users' satisfaction, requires both patience and discipline by the top management and the staff of distribution utilities to admit their level of knowledge in the past regarding power quality issues, and what the existing problems are they still face. This would help them to learn better how to avoid these obstacles, by raising their awareness of power quality [7].

A study conducted by Milanovic and Negnevitsky in Croatia stressed that the expected benefits of PQP implementation would make significant contributions to customer satisfaction [48]. They suggested that this level of customer satisfaction could be used as part of the process to identify the level of PQ issues, and where improvements could be made to increase this satisfaction. Barnard and Van Voorhis found that the main benefits of PQP implementation are increasing end users' awareness, increasing end users' satisfaction and improving power quality performance [2]. Labricciosa in his study in Canada stated that successful PQP implementation will result in reducing end users' complaints, and solving PQ disturbances [31].

A study by Aniruddh in the USA found that one of the main benefits of implementing PQP was to; provide PQ diagnosis systems and databases offering adequate data for end users to tackle PQ disturbances themselves, as their awareness level increased [38]. Janjic, Stajic and Radovic stated that when PQP was implemented successfully the distribution utilities gained the benefits of strategic planning by taking appropriate action, and making adequate preparations to introduce effective changes in the distribution systems regarding PQ; and satisfying their customers [49].

A study by Ronghua and Suan in Singapore found that end users' satisfaction and reducing PQ cost are the most valuable benefits of implementing PQP [24]. A study by Qureshi and Paracha in Pakistan found that the great benefits of PQP implementation are reducing the pressure of demand, improving network performance, increasing top management awareness, and developing the distribution systems for future requirements [36].

A survey by Salam and Nasri in Egypt found that one of the benefits of PQP implementation is to, increase customer satisfaction, raising the level of employee skills and awareness to tackle PQ issues [50].

A study by Gul in Turkey found that the most valuable benefits of implementing PQP are measuring PQ disturbances, increasing PQ training courses, providing enough knowledge, widening employees' experience and skills, educating end users and engineers and reducing the huge losses for end users and utilities [35].

The benefits of PQP revealed by the above researchers and studies can be summarized as: increasing end users' awareness and , satisfaction, improving power quality performance, reducing end users' complaints, monitoring and measuring PQ disturbances, providing PQ diagnosis systems and databases, providing strategic planning, reducing PQ cost, improving network performance, increasing top management awareness, raising the level of employee skills, experience, knowledge and awareness, increasing PQ training courses, and reducing the huge losses for end users and utilities.

In Libyan distribution networks, empirical research is required to categorize and underline the barriers and benefits of PQP in the context of a distribution utility, which has not implemented power quality programs in the last two decades. The knowledge and results obtained from this study will guide Libyan distribution networks implementing PQP, including all departments and staff, who are directly responsible for remedying power quality disturbances, in tackling any power quality problems by setting up clear and long-term strategies, with crucial objectives and serious barriers. Therefore, the implementation of power quality program requires great attention from the top management to help the distribution networks to achieve their goal of offering and providing a power quality program in practice [51].

5 RESEARCH METHOD AND SURVEY INSTRUMENT

The above literature review helps the researcher to understand the different barriers to PQP implementation and the expected benefits of PQP. Next, a survey questionnaire and interviews were conducted in the Libyan distribution systems. The questionnaire was designed in two main parts and, followed by 44 face-to-face interviews.

5.1 Part A

Respondents were asked to define how far any of the 16 PQP potential barriers (BA) cause current difficulties in implementing power quality program in Libyan distribution systems. The 16 PQP barriers are listed in table 3. All factors were designed in a five-point Likert scale format (1=not

applicable; 2= very low extent; 3= low extent; 4= moderate; 5= high extent).

TABLE 3
LIST OF POWER QUALITY PROGRAM BARRIERS

PQP Barriers	
BA1	lack of staff awareness, skills and experience
BA2	lack of end users awareness
BA3	lack of customer cooperation
BA4	lack of long-term strategy and planning
BA5	lack of top management commitment
BA6	lack of network designing
BA7	lack of distribution networks infrastructure
BA8	lack of conducting research and studies
BA9	lack of top management responsibility
BA10	lack of training courses, and support
BA11	lack of financial resources
BA12	lack of enough incentives
BA13	lack of PQ measurement
BA14	lack of PQ consultants
BA15	lack of PQ standards
BA16	lack of PQ monitoring and database

5.2 Part B

Respondents were asked to judge how far one of 11 PQP possible benefits (BN) would be achieved by implementing PQP within Libyan distribution systems. The 11 PQP expected benefits are listed in table 4. All factors were designed in five-point Likert scale format (1= not sure; 2=negative; 3= moderate; 4= positive; 5= very positive).

TABLE 4
LIST OF POWER QUALITY PROGRAM BENEFITS

PQP Benefits	
BN1	Increasing the end users awareness
BN2	Increasing the end users satisfaction
BN3	Improving PQ performance
BN4	Reducing the end users complaints
BN5	Monitor & Measuring PQ disturbances
BN6	providing PQ diagnosis system and database
BN7	Reducing the huge losses of PQ cost
BN8	Increasing the top management awareness
BN9	Increasing the employee skills and awareness
BN10	Increasing PQ training courses
BN11	Providing strategic planning

The questionnaire was sent to head managers, middle managers, engineers, technicians and employees, with total number of 540 copies and it conducted in April-June 2009. Of 540 copies, 441 copies were returned, of which 397 were appropriate for data analysis, giving a response rate of 81%. The data were analyzed by using Statistical Package for Social Science (SPSS) software, version 15.0.1.1.

In addition, 44 interviewees participated in this study to investigate why there were barriers to PQP implementation.

The interviewees consisted of head managers, middle managers, engineers, technicians and employees from four departments, mainly those dealing directly with power quality issues. These were Planning, Training, Distribution, and Customer departments in LDNs. After the interviews conducted the data were transcribed and coded by using NVivo 9 [52], [53].

6 RESULTS AND DISCUSSION

Data gathered by the questionnaire from the distribution system respondents were checked in terms of accuracy, outliers and, normality; then analyzed using (SPSS) software. Table 5 shows the type of distribution networks along with the categories of end users involved in the study. Large distribution networks were considered to have more categories of end users; the western distribution network (WDN1), southern-west distribution network (SWDN2) and eastern distribution network (EDN4); whereas small distribution networks had 1 to 2 categories of end users; the central distribution network (CDN3) and southern-east distribution network (SEDN5).

TABLE 5
TYPES OF DISTRIBUTION NETWORKS

Distribution Networks	Residential	Commercial	Industrial	Agricultural
WDN1	√	√	√	
SDN 2	√		√	√
CDN3	√	√		
EDN 4	√		√	√
SDN 5	√			√

The data were measured in order to evaluate the correlations between the barriers to PQP; therefore factor analysis was performed. The Kaiser-Meyer-Oklin (KMO) measure of sampling Adequacy value was 0.82, which exceeds the recommended value of 0.6 [54] and the Bartlett's Test of Sphericity was also highly significant (Chi-Square = 4847.51 with 561 degrees of freedom, at $p < 0.001$), reaching statistical significance in the correlation matrix. This implies that the factor analysis of 16 factors of PQP barriers was appropriate and confirms that all the items were statistically significant, which are judged to be an excellent validation of factor analysis.

The reliability test of Cronbach's α for all factors in parts A and B of questionnaire is 0.82. Cronbach's α above the cited minimums of 0.70 [55] is considered to be high and acceptable alpha, giving an evidence that the total Cronbach's alpha was judged to be reliable for the questionnaire.

Table 6 summarizes the Varimax - rotated factor matrix, which accounted for about 64 % of the total variation. The correlation matrix disclosed the presence of many items < 0.5 and items higher were considered to be important. Questions BA 1-4 belong to factor 1 and can be categorized

under 'lack of awareness', whereas questions BA 5-9, belong to factor 2 and are categorized as 'lack of top management attention'. Questions BA 10-12 belong to factor 3 and pertain to 'lack of resources' and finally questions BA 13-16 belong to factor 4, dealing with 'lack of power quality involvement'.

6.1 Part A

Table 6 shows that in the ANOVA test, out of 16 barriers, 12 were statistically significant different at the P value <0.05. The significant barriers are BA1, lack of staff awareness, skills and experience, BA2, lack of end users awareness, BA4, lack of long-term strategy and planning, BA5, lack of top management commitment, BA6, lack of network designing, BA7, lack of distribution networks infrastructure, BA9, lack of top management responsibility, BA10 lack of training courses and support, BA11, lack of financial resources, BA13, lack of PQ measurement, BA14, lack of PQ consultants, BA15, lack of PQ standards, and BA16, lack of PQ monitoring and database.

TABLE 6
VARIMAX ROTATED FACTOR MATRIX

Items	Factor 1	Factor 2	Factor 3	Factor 4	Sig
BA1	0.797				0.035
BA2	0.731				0.033
BA3	0.699				0.337
BA4	0.666				0.036
BA5		0.801			0.044
BA6		0.754			0.049
BA7		0.676			0.021
BA8		0.641			0.447
BA9		0.623			0.043
BA10			0.837		0.022
BA11			0.787		0.044
BA12			0.755		0.242
BA13				0.766	0.031
BA14				0.711	0.041
BA15				0.701	0.029
BA16				0.671	0.128

In addition, a post hoc Least Significance Difference (LSD) test was carried for these twelve barriers. The test found that large distribution networks WDN1, SDN2 and EDN4 faced some particular barriers compared to other small distribution networks in LDNs. SDN2 faces three factors; F1, lack of awareness, F2, lack of top management attention, and F4, lack of PQ involvement, whereas WDN1 and EDN4 face F1, lack of awareness, F4, lack of PQ involvement and F3, lack of resources. As a result, it can be referred that Libya's distribution systems have so far struggled to implement PQP effectively.

6.2 Part B

This part of the questionnaire shows the analysis of the mean level of PQP benefits within LDNs. The response scale of the survey was divided into three levels of outcome, where

(1.51 to ≤ 250 was Negative, 2.51 to ≤ 3.50, moderate and, 3.51 to ≤ 5 Positive). Table 7 presents the overall results of PQP benefits, which would have a positive impact on increasing end users awareness, increasing their satisfaction, improving PQ performance, reducing end users' complaints, monitoring and measuring PQ disturbances, providing PQ diagnosis systems and databases, reducing the huge losses through PQ costs, increasing top management awareness, increasing the employees' skills and awareness, increasing PQ training courses and providing strategic planning in LDNs.

TABLE 7
BENEFITS OF SURVEY RESULTS

No	DN1	DN2	DN3	DN4	DN5	Overall
BN1	3.84	3.96	3.45	3.27	3.8	3.66
BN2	3.91	3.56	3.54	3.73	3.53	3.65
BN3	3.65	3.68	3.54	3.64	3.4	3.58
BN4	3.51	3.52	3.68	3.51	3.47	3.53
BN5	3.48	3.48	3.82	3.53	3.33	3.52
BN6	3.73	3.56	3.67	3.49	3.46	3.58
BN7	3.52	3.48	3.49	3.55	3.66	3.54
BN8	3.76	3.88	3.82	3.77	3.93	3.83
BN9	4.25	3.31	3.75	3.53	3.48	3.66
BN10	3.43	3.68	3.73	3.25	3.52	3.52
BN11	3.48	3.66	3.61	3.52	3.56	3.56

7 INTERVIEW RESULTS

Table 8 shows the twelve difficulties discussed in the interviews, which are similar to what were obtained from the questionnaire. These results indicate that LDNs have not implemented PQP. It showed that the top management has not paid enough attention, support, commitment and responsibility to setting up long-term strategies to implement PQP. Therefore, LDNs have lost LD 464 million annually due to poor power quality and the failure to implement PQP [56].

TABLE 8
BARRIERS TO POWER QUALITY PROGRAM IMPLEMENTATION FROM ANALYSIS OF THE INTERVIEWS

Barriers	Head Managers	Middle Managers	Engineers	Technicians
BA1	4.5%	69.85%	12.64%	13.01%
BA2	2.85%	56.26%	20.38%	20.51%
BA4	3.9%	60%	17.18%	18.92%
BA5	7.56%	56.68%	17.91%	17.84%
BA6	2.32%	71.44%	16.12%	10.12%
BA7	17.64%	50.1%	16.93%	15.33%
BA9	6.12%	76.75%	17.13%	0%
BA10	16.53%	44.35%	3.72%	35.4%
BA11	2.53%	58.26%	7.5%	31.71%
BA13	0%	95.27%	0%	4.73%
BA15	3.08%	83.28%	2.91%	10.73%
BA16	8.81%	64.18%	16.67%	10.34%

Moreover, most of members of staff involved in implementing PQDs are middle managers, 52.4% of who held of high diploma qualifications, which is considered the minimum educational level. This means that they are not highly knowledgeable and aware enough to cope with the current severe level of power quality as well; moreover, this level of education would not enable them to understand and participate in implementing PQP. Almost 38% of engineers and technicians have between 6 and 15 years of experience, but lack awareness and skills. They should be better taught and trained before they can deal with PQP implementation.

8 PROPOSED PQP FRAMEWORK

Multivariable Linear Regression (MVLRL) was conducted to identify which factors have significant impact on PQP implementation [57]. An acceptable model was developed on the basis of these factors. It is clear that all these factors are significantly correlated, since all p values are less <0.05 and are substantially affected by the lack of awareness of the implementation of PQP in Libyan distribution networks as shown in Fig.1.

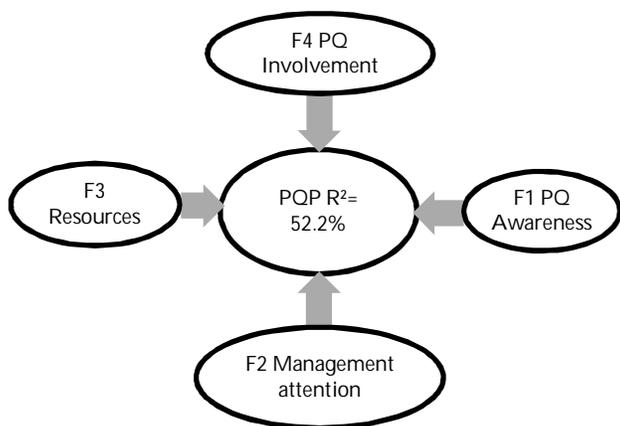


Fig. 1. Power Quality Program Model for LDNs

Table 9 shows the value of R² as 52.2% for this model, which indicates how much of the variability in the outcome is explained by the predictors. This also indicates that the validity of this model is very good. Consequently, this model can be accepted and applied for LDN to implement PQP, since all the predictors increase by one unit (see β value). It also indicates that the two factors most highly affected by lack of PQ awareness are F2 (β=34.5%) and F3 (β=31.6%). As a result, the regression analysis shows that the linear relationship between the outcomes, which is PQP, is explained by the model and predictor factors.

TABLE 9

REGRESSION RESULTS OF POWER QUALITY PROGRAM FACTORS

Scale	β	Std .Error	t	P	R ²	Cronbach's alpha
F1	0.202	0.031	4.538	<.001	0.522	0.811
F2	0.345	0.041	7.573	<.001		0.841
F3	0.316	0.029	8.097	<.001		0.806
F4	0.171	0.028	4.427	<.001		0.851

9 CONCLUSION

This study is the first to investigate the barriers and benefits of PQP within Libyan distribution systems. It contributes by providing an insight into the overall efforts needed to implement PQP implementation and the main reasons underlying its failure. It is also the first to explore the expected benefits, to be gained from implementing PQP. The findings will be applied to build a PQP framework guideline to be implemented in LDNs. Four main factors of PQP barriers were determined from this study, namely; lack of awareness (lack of staff awareness, skills and experience, lack of end users' awareness, lack of customer cooperation, lack of long-term strategy and planning); lack of top management attention (lack of top management commitment, lack of network designing, lack of infrastructure for distribution networks, lack of continuing research and study, lack of top management responsibility); lack of resources (lack of training courses and support, lack of financial resources, lack of enough incentives); lack of power quality involvement (lack of PQ measurement, lack of PQ consultants, lack of PQ standards, lack of PQ databases).

The large distribution networks WDN1, SDN2 and EDN4 faced some particular barriers, unlike the smaller distribution networks in LDNs. SDN2 faces three factors F1, lack of awareness, F2, lack of top management attention, and F4, lack of PQ involvement; whereas WDN1 and EDN4 face F1, lack of awareness, F4 lack of PQ involvement and F3 lack of resources. The result of this is that Libya's distribution systems have struggled so far to implement PQP effectively. In general, the finding shows that LDNs suffer the four factors of PQP barriers. These four factors appeared in USA, European, India, Malaysia, Latin America, Brazil, Germany, Pakistan, Austria, France, Italy, Poland, Portugal, Slovenia, Spain and UK.

The implementation of PQP plays a significant role in improving power quality issues. The purpose of implementing PQP is associated with completing and developing systems to achieve the strategy's objectives set by all departments. Therefore, this study reveals poor implementation of PQP in LDNs, because they are not moving from the suggested strategies to realistic performances. According to qualitative analysis, this gap will continue if PQP is not implemented. Therefore, one of the main challenges in implementing PQP is to link all the difficulties with both its objectives and strategies. Hence, the implementation difficulties should be

regularly assessed to identify the hidden reasons associated and causing poor implementation. Thus, without adequate knowledge, awareness, planning, designing, preparation, training, power quality standards, clear strategy, and most important the support of top management for this program, power quality disturbances will never end and their severity will affect all consumers.

In response to this, a POP for LDNs was found to have a positive impact on increasing end users' awareness, and satisfaction, improving PQ performance, reducing end users' complaints, monitoring and measuring PQ disturbances, providing a PQ diagnosis system and database, reducing the huge losses through PQ cost, increasing top management awareness, increasing employees' skills and awareness, increasing PQ training courses and providing strategic planning. They are needed because LDNs have not yet implemented POP due to the failure to establish power quality departments.

REFERENCES

- [1] B. E. Kushare, A. A. Ghatol, "Power quality characteristics of industrial electric distribution systems," Institute of Engineering Education & Research, Nashik, India, 2008.
- [2] A. Barnard, A. Van Voorhis. Investigation utility-based power quality programs: Domestic and international perspectives. EPRI. USA. 2000.
- [3] J. V. Milanovic and M. Negnevitsky, "Power quality problems and solutions: Current understanding," in *Harmonics and Quality of Power*, 1998. Proceedings. 8th International Conference on, 1998, pp. 30-35 vol.1.
- [4] H. M. S. C. Herath. Power quality data management and reporting methodologies. [University of Wollongong]. 2008.
- [5] M. R. Wadi, M. F. Bara, O. Carlson and K. M. Elarroudi, "Voltage stability analysis for the south-west Libyan electrical power system: Problem simulation and analysis," in *Universities Power Engineering Conference (UPEC)*, 2009 Proceedings of the 44th International, 2009, pp. 1-5.
- [6] G. Nicholson, "Investigation of Data Reporting Techniques and analysis of Continuous Power Quality Data in the Vector Distribution Networks," 2006.
- [7] The National Energy Technology Laboratory, "A Systems View of the Modern Grid: Provides Power Quality For 21st Century Needs," vol. 2, pp. 1-24, 2007.
- [8] M. A. Hannan, A. Hussain, A. Mohamed and R. A. Begum, "Methodological Framework for Industrial Flicker Severity Assessment and Awareness," *Australian Journal of Basic & Applied Sciences*, vol. 4, pp. 5578-5583, 11, 2010.
- [9] D. Kottick. *Electrical power quality & utilization magazine, power quality monitoring System—Voltage dips, short interruptions and flicker*. 3(2), pp. 2. 2008.
- [10] General Electrical Company of Libya, "Annual report," GECOL, Tripoli, Libya, 2009.
- [11] A. H. Hadod. *Libyan electrical power sector*. GECOL. Tripoli, Libya. 2009.
- [12] B. T. M. Elobadi, "Power quality report : Voltage profile," GECOL, Tripoli, Libya, 2009.
- [13] M. A. Mousa, I. M. Saleh Ibrahim and I. M. Molokhia, "Comparative study in supplying electrical energy to small remote loads in Libya," *Renewable Energy*, vol. 14, pp. 135-140, 8, 1998.
- [14] D.F. Warne. *Newness Electrical Power Engineer's Handbook* (2nd ed.) 2005.
- [15] A. Lima, C. Barreiero, M. DeMarco, J. Martins, R. Roncolato, and N. Santos, "Cost of Power Quality Problems in Large Industrial Customers," *Proceedings of POA '93/PECON IV Conference and Exhibition, Power Quality: Issues and Opportunities, Power Electronics: Solutions and Responses*, vol. 1, 1993.
- [16] R. Caldon, M. Fauri, L. Fellin, "Voltage Sag Effects on Continuous Industrial Processes: Desensitizing Study for Textile Manufacture," *Proceedings of the Second International Conference on Power Quality: End-use Applications and Perspectives*, 1992.
- [17] H. Morosini, C. Burns, C.A. Warren, T. Dack, J. Burke, T. Short, "Cost of momentary interruptions and voltage sags to an industrial customer," *IEEE Summer Power Meeting*, 1995.
- [18] C. R. Cleavelin, "Power Quality: A Semiconductor Manufacturer's Perspective," Presented at PQ for the Semiconductor Fabrication Industry, vol. 2, 1997.
- [19] V. Wagner, T. Grebe, R. Kretschmann, L. Morgan, A. Price, "Power System Compatibility with Industrial Process Equipment," A Panel Discussion," *Conference Record of the 1994 IEEE Industry Applications Society, 29th Annual Meeting*, vol. 3, 1994.
- [20] J. M. R. Targosz, "Pan-european power quality survey," in *9th International Conference on Electrical Power Quality and Utilisation, EPQU 2007*, 2007, pp. 1-6.
- [21] B. H. S. McNulty. *Power quality problems and renewable energy solutions, analysis of the prospects for renewable PQ solutions in massachusetts*. Massachusetts Renewable Energy Trust. USA. 2002.
- [22] D. Lineweber, S. R. McNulty, "The cost of power disturbances to industrial & digital economy companies," *EPRI IntelliGrid Initiative (A Primen report from EPRI & CEIDS, USA, Tech. Rep. 1006274*, 2001.
- [23] R. Langley, A. Mansoor, E. R. Collins Jr. and R. L. Morgan, "Voltage sag ride-through testing of adjustable speed drives using a controllable dynamic dynamometer," in *Harmonics and Quality of Power, 1998. Proceedings. 8th International Conference on*, 1998, pp. 566-571 vol.1.
- [24] G. R. Ronghua, H. C. Suan, "A customer - oriented approach in managing power quality in Singapore," *International Power Quality Conference IPQC 2002*, vol. 2, pp. 188- 192, 2002.
- [25] F. Carlsson, B. Widell and C. Sadarangani, "Ride-through investigations for a hot rolling mill process," in *Power System Technology, 2000. Proceedings. PowerCon 2000. International Conference on*, 2000, pp. 1605-1608 vol.3.
- [26] S. M. D. Lineweber, "The cost of power disturbances to industrial & digital economy companies," *EPRI's Consortium for Electric Infrastructure for a Digital Society (CEIDS), USA, Tech. Rep. 108829*, 2001.
- [27] H. De. Keulenaer, "The Hidden Cost of poor Power Quality," pp. Leonardo Energy, European Copper Institute, 2003.
- [28] C. C. Pearson, V. Uthayanan, "THE BSRIA POWER QUALITY GUIDE," vol. Application Guide AG 2, pp. 1-3, 2000.
- [29] B.W. Kenedy, *Power Quality Primer*. USA: McGraw-Hill, 2000.
- [30] J. Arrillaga, M. H. J. Bollen and N. R. Watson, "Power quality following deregulation," *Proceedings of the IEEE*, vol. 88, pp. 246-261, 2000.
- [31] I. N. Labricciosa. *A case study of power quality enhancement in an electrical distribution system*. 1996.
- [32] R.C. Dugan, M.F. Mcgrnanahan, S. Santoso, H.W. Beaty. *Electrical Power System Quality (Second ed.)* 1996.
- [33] I. Chung, D. Won, J. Kim, S. Ahn and S. Moon, "Development of a network-based power quality diagnosis system," *Electr. Power Syst. Res.*, vol. 77, pp. 1086-1094, 6, 2007.
- [34] J. G. D. L.A. Schienbein. *Distributed energy resources, power quality and reliability –Background*. PACIFIC NORTHWEST NATIONAL LABORATORY. USA. 2002.

- [35] O. GUL, "An Assessment of Power Quality and Electricity Consumer's Rights in Restructured Electricity Market in Turkey," Istanbul Technical University, Istanbul, vol. XIV, pp. 29-34, 2008.
- [36] Z. J. Paracha, S. A. Qureshi, "Application OF Power Quality Management Techniques and Methods in Power System of WAPDA PAKISTAN," vol. 4, pp. 1813-419, 2007.
- [37] Ph. Ferracci, "Power quality," Cahier Technique Schneider Electric, UK, Tech. Rep. 199, 2001.
- [38] G. D. N. Aniruddh. Technical report on power quality management. Electrical and Electronics ENGG. USA. 2003.
- [39] J. Meyer, P. Schegner, G. Winkler, M. Muhlwitz and L. Schulze, "Efficient method for power quality surveying in distribution networks," in Electricity Distribution, 2005. CIRED 2005. 18th International Conference and Exhibition on, 2005, pp. 1-4.
- [40] C. M. D. Dorr. Power quality for satisfied commercial and residential customers field test plan: Monitoring residential power quality. EPRI. USA. 2000.
- [41] W. T. J. Hulshorst, E. L. M. Smeets, and J. A. Wolse, "Premium power quality contracts and labeling," Work package 2 of the Quality of Supply and Regulation project, Netherlands, Tech. Rep. 30620162-Consulting 07-0401, 2007.
- [42] U. K. Deshpande, P. A. Chitre, "Total power quality management in a large distribution system: A case study national seminar on "Total power quality management", " Mumbai, India, 2009.
- [43] N. Rajmani, "ELECTRIC POWER QUALITY: SOME TECHNICAL AND COST PERSPECTIVES " Global Energy Consulting Engineers, INDIA, Tech. Rep. 265, 2008.
- [44] S. A. Qureshi, F. Mahmood, Z. J. Paracha, S. A. Kashif, "POWER QUALITY DISTURBANCES IN POWER SYSTEMS AND COMPARISON OF DEVELOPED COUNTRIES STANDARDS," Electrical Engineering Department, UET Lahore, Pakistan, vol. 4, 2007.
- [45] W. Moncrief, J. Dougherty, L. Richardson, K. Craven, "Power quality applications guide for architects and engineers," EPRICSG, Palo Alto, CA, USA, Tech. Rep. 113874, 1999.
- [46] B. C. A. Abreu, "Power quality site survey process for regulated electricity market," in Transmission & Distribution Conference and Exposition: Latin America, 2006. TDC '06. IEEE/PES, 2006, pp. 1-6.
- [47] R. J. R. Gomes, D. O. C. Brasil and J. R. de Medeiros, "Power quality management issues over the brazilian transmission system," in Harmonics and Quality of Power, 2002. 10th International Conference on, 2002, pp. 27-32 vol.1.
- [48] A. Eberhard, Ed., POWER QUALITY. Croatia: InTech Janeza Trdine 9, 51000 Rijeka, 2011.
- [49] A. JANJIĆ, Z. STAJIĆ, I. RADOVIĆ, "Power Quality Issues in Smart Grid Environment: Serbian Case Studies," Serbian Ministry of Education and Science (Project III44006) SERBIA, pp. 2-3, 2011.
- [50] G. A. A. Salam and S. M. Nasri, "Survey of power quality problems in industrial zones in egypt," in Electricity Distribution, 2005. CIRED 2005. 18th International Conference and Exhibition on, 2005, pp. 1-2.
- [51] W. E. Kazibwe, R. J. Ringlee, G. W. Woodzell and H. M. Sendaula, "Power quality: a review," Computer Applications in Power, IEEE, vol. 3, pp. 39-42, 1990.
- [52] D. Silverman, Ed., Interpreting Qualitative Data. Thousand Oaks, California: Sage, 2006.
- [53] G.R. GIBBS, Ed., Qualitative Data Analysis "Explorations with NVivo". New York: Penn Plaza, 2002.
- [54] A. P. Field, Ed., Discovering Statistics using SPSS: (and Sex, Drugs and Rock'n'Roll). London; Thousand Oaks: SAGE Publications, 2005.
- [55] J. C. Nunnally, I. H. Bemstein, Psychometric Theory. New York: McGraw-Hill, 1994.
- [56] B. A. Hamilton, "Strategic plan corporate objectives; business, functional unit plans, financial model and lessons learned," GECOL, Tripoli, Libya, 2006.
- [57] H. M. S. Chandana Herath, V. J. Gosbell, S. Perera and D. Stirling, "Power quality survey factor analysis using multivariable linear regression (MVLR)," in Harmonics and Quality of Power, 2008. ICHQP 2008. 13th International Conference on, 2008, pp. 1-5.

AUTHORS' INFORMATION

Saad S. Sultan has received his BSc degree in Communication from Ibn-AL Hytham High Centre for Technological Education and Scientific Research, Tripoli, Libya and the MSc degree in Electrical Engineering with Power Electronics from Bradford University, UK, in 2007. He has joint Brunel University in 2009 to carry out his PhD research in the area of Power Quality Issues. He is a student member of IET in Power Electronics System Network in field of Power Quality.

Mohamed K. Darwish (BSc, PhD, IEE Member 89-CEng 90, MIEEE). Dr Darwish has received the BSc degree in Electrical Engineering from Helwan University, Cairo, Egypt, and the PhD degree from Brunel University in 1978 and 1987 respectively. Dr Darwish is currently a Senior Lecturer in Power Electronics. He has over 20 years experience of research and teaching in the area of Power Electronics. His research interests include active filtering techniques, UPS systems and Power Quality issues. Dr. Darwish has published over 100 papers in leading journals and conferences.

Performance Analysis of Subspace Methods Used in Blind Channel Identification

T.Priyadarsini

Abstract — A subspace-based method is proposed for estimating the channel responses of single-input-multiple-output (SIMO) Orthogonal Frequency Division Multiplexing (OFDM) system. Our technique relies on minimum noise subspace (MNS) decomposition to obtain noise subspace in a parallel structure from a set of pairs (combinations) of system outputs that form a properly connected sequence (PCS). The developed MNS-OFDM estimator is more efficient in computation than subspace (SS)- OFDM estimator, although the former is less robust to noise than the later. To maximise the MNS-OFDM estimator performance, a symmetric version of MNS is implemented. We present simulation results demonstrating the channel identification performance of the corresponding OFDM-based SIMO systems employ cyclic prefixing approach.

IndexTerms — Blind Channel Identification, Equalisation, MNS, OFDM

1 INTRODUCTION

OFDM is a multi-carrier digital modulation technique that facilitates the transmission of high data rates with a limited bandwidth. It is an effective technique for several applications such Digital Audio Broadcasting (DAB) and terrestrial Digital Video Broadcasting (DVB). In addition, OFDM forms the basis for the physical layer in upcoming standards for broadband Wireless Local Area Network (WLAN) , i.e. ESTI-BRAN HiperLAN/2 , IEEE 802.11a and Multimedia Mobile Access Communication Systems (MMAC) and for Fourth Generation (4G) broadband wireless systems that will perform multimedia transmission to mobiles and portable personal communications devices, i.e. European MEMO project and for IEEE 802.16.

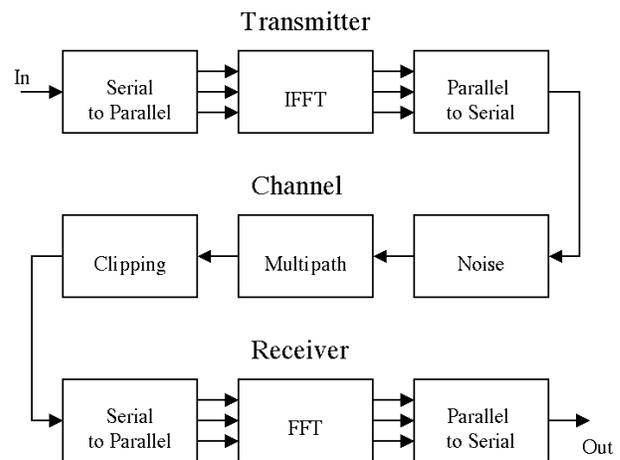


Fig 1- Block Diagram Of OFDM

The transmitter first converts the input data from a serial stream to parallel sets. Each set of data contains one symbol, S_i , for each subcarrier. For example, a set of four data would be $[S_0 S_1 S_2 S_3]$.

Before performing the Inverse Fast Fourier Transform (IFFT), this example data set is arranged on the horizontal axis in the frequency domain as shown in Figure 2. This symmetrical arrangement about the vertical axis is necessary for using the IFFT to manipulate this data.

- T.Priyadarsini, M.E., Asst.Prof/ ECE, V.R.S college of Engineering and Technology, Villupuram, TamilNadu, India. E-mail: priyadarsinithiruvengadam@gmail.com

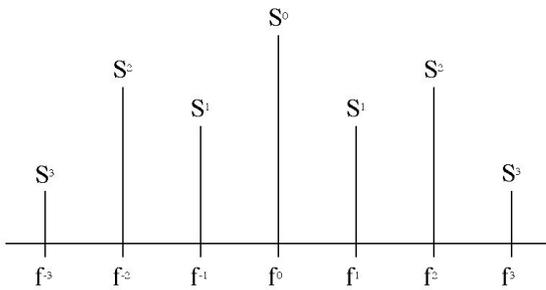


Fig 2- Frequency Domain Distribution Of Symbols

An inverse Fourier transform converts the frequency domain data set into samples of the corresponding time domain representation of this data. Specifically, the IFFT is useful for OFDM because it generates samples of a waveform with orthogonal frequency components. Then, the parallel to serial block creates the OFDM signal by sequentially outputting the time domain samples.

The channel simulation will allow examination of the effects of noise, multipath, and clipping. By adding random data to the transmitted signal, simple noise can be simulated. Multipath simulation involves adding attenuated and delayed copies of the transmitted signal to the original. This simulates the problem in wireless communication when the signal propagates on many paths. For example, a receiver may see a signal via a direct path as well as a path that bounces off a building. Finally, clipping simulates the problem of amplifier saturation. This addresses a practical implementation problem in OFDM where the peak to average power ratio is high.

The receiver performs the inverse of the transmitter. First, the OFDM data are split from a serial stream into parallel sets. The Fast Fourier Transform (FFT) converts the time domain samples back into a frequency domain representation. The magnitudes of the frequency components correspond to the original data. Finally, the parallel to serial block converts this parallel data into a serial stream to recover the original input data.

Due to increase in the normalised delay spread, multipath fading becomes a major concern as systems with high data rate are more liable to intersymbol interference (ISI). Classically, ISI is eliminated by employing a cyclically extended time domain guard interval (GI). Thus, each OFDM symbol is preceded by a periodic extension of the symbol itself. This GI is also known as cyclic prefix (CP) and the system CP-OFDM. Recently, zero-padding OFDM (ZP-OFDM), which pre-pends each OFDM symbol with zeros rather than replicating the last few samples, has been proposed. ZP-OFDM not only has all the advantages of the CP-OFDM, but also guarantees symbol recovery and ensures finite impulse response (FIR) equalisation. However, the implementation of a ZP-OFDM system involves transmitter modifications and complicates the equalizer.

To maximise the performance advantage of OFDM system, reliable identification of Single Input Multiple Output (SIMO) channels is desired. Currently, the channel identification and equalisation technique used requires a major fraction of the channel capacity to send a training sequence over the channels. There are practical situations where it is not feasible to utilize a training sequence such as in fast varying channels. To save this fraction of channel capacity, blind identification is an attractive approach. Using the blind channel identification techniques, the OFDM-based SIMO receiver can identify the channel characteristics and equalises the channel all based on the received signal, and no training sequence is needed, which hence saves the channel capacity.

Blind identification and equalisation of SIMO channels have been a very active area of research during the past few years. Among the various known algorithms, Second Order Statistics (SOS)-based algorithms are the most attractive due to their special properties. It was, for a while, believed that the subspace (SS)-based method was the only key to the surprising success among the existing SOS-based techniques. The SS-based method applies the Multiple Signal Classification (MUSIC) concept to a relation between the channel impulse responses and the noise subspace associated with a covariance matrix of the system output.

One of the important advantages of SS-based method is its deterministic property. That is, the channel parameters can be recovered perfectly in the absence of noise, using only a finite set of data samples, without any statistical assumptions over the input data. More recently, the use of the SS-based method has been suggested to accomplish blind SIMO channel identification in OFDM systems. Despite their high identification efficiency, SS-based methods are computationally very intensive, which may be unrealistic or too costly to implement in real time, especially for large sensor array systems. The main reason is that they require non-parallelisable eigen-value-decomposition (EVD) of a large dimensional matrix to extract (estimate) the noise or signal subspace.

In this paper, using a minimum noise subspace (MNS) decomposition concept, we introduce several techniques for blind identification and equalisation of OFDM-based SIMO systems. Our techniques compute the noise subspace via a set of noise vectors (basis of the noise subspace) that can be computed in parallel from a set of pairs (combinations) of system output, without using reference or pilot symbols. Therefore, an EVD for smaller covariance matrices is required to extract noise subspace. Ideally, this approach, which relies on the known structure of the received OFDM symbols, provides a perfect channel estimate in the absence of noise. It is believed to have inspired all the subsequent developments which have taken place to accomplish unknown parameter identifications in a wide range array signal processing applications. Furthermore, the developed techniques significantly reduce receiver complexity in wireless broadband multi-antenna systems.

2 PROPOSED WORK

In this section, OFDM-based SIMO is introduced by using CP and ZP techniques.

2.1 Standard CP-OFDM System

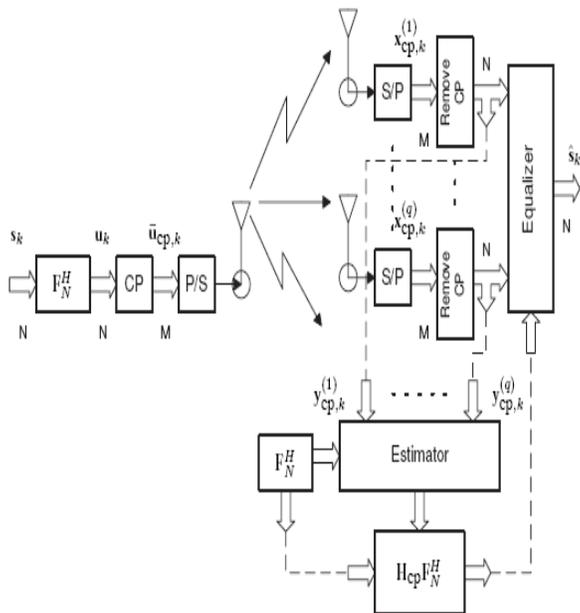


Fig. 3 - CP-OFDM system: transmitter and receiver.

Figure 3 depicts the baseband discrete-time block equivalent model of a standard CP-OFDM system. The transmitted symbols are parsed into blocks of size N : $\mathbf{s}_k = [s_k(0), s_k(1), \dots, s_k(N-1)]^T$ where $k = 0, 1, 2, \dots, K-1$. The elements of \mathbf{s}_k are considered to be independent and identically distributed (i.i.d). We regard these elements to be in the frequency domain. The symbol block \mathbf{s}_k is then modulated and converted into time domain using the IFFT matrix \mathbf{F}_N^H , where \mathbf{F}_N has entries

$$f_{n,d} = \frac{1}{N} \exp\left(\frac{j2\pi nd}{N}\right)$$

and $d, n = 0, \dots, N-1$. The data vector $\mathbf{u}_k = \mathbf{F}_N^H \mathbf{s}_k = [u_k(0), u_k(1), \dots, u_k(N-1)]^T$ is then appended with a CP of length L , resulting in a size $M = N + L$ signal vector: $\mathbf{u}_{cp,k} = \mathbf{T}_{cp} \mathbf{u}_k = [u_k(N-L), \dots, u_k(N-1), u_k(0), \dots, u_k(N-1)]^T$. We consider \mathbf{T}_{cp} is a concatenation of the last L rows of an $N \times N$ identity matrix \mathbf{I}_N (that we denote as \mathbf{I}_{cp}) and the identity matrix itself \mathbf{I}_N , i.e., $\mathbf{T}_{cp} = [\mathbf{I}_{cp}^T, \mathbf{I}_N^T]^T$. CP makes the OFDM appear periodic over the time span of interest. The channel response is denoted by $h^{(r)}(l)$ where $l = 0, 1, \dots$,

$L^{(r)}$, and $r = 1, 2, \dots, q$. To avoid ISI, as indicated previously, the CP length L is selected to be equal to or greater than the channel order, i.e., $L^{(r)} \leq L$. We consider the upper bound of the SIMO channel order $L^{(r)}$ as a CP length L . The received k -th block at r -th output for $n = 0, 1, \dots, M-1$, is given by

$$x_{cp,k}^{(r)}(n) = \sum_{l=0}^L h^{(r)}(l) u_{cp,k}(n-l) + v_{cp,k}^{(r)}(n) \quad (1)$$

where $u_{cp,k}(n-l)$ and the AWGN, $v_{cp,k}^{(r)}(n)$, is assumed to be mutually uncorrelated and stationary. Using the following notations

$$\begin{aligned} \mathbf{x}_{cp,k}(n) &= [x_{cp,k}^{(1)}(n), x_{cp,k}^{(2)}(n), \dots, x_{cp,k}^{(q)}(n)]^T \\ \mathbf{v}_{cp,k}(n) &= [v_{cp,k}^{(1)}(n), v_{cp,k}^{(2)}(n), \dots, v_{cp,k}^{(q)}(n)]^T \\ \mathbf{h}(l) &= [h^{(1)}(l), h^{(2)}(l), \dots, h^{(q)}(l)]^T \end{aligned} \quad (2)$$

we can rewrite the input-output relation (1) in vector matrix as

$$\mathbf{x}_{cp,k}(n) = \sum_{l=0}^L \mathbf{h}(l) u_{cp,k}(n-l) + \mathbf{v}_{cp,k}(n). \quad (3)$$

2.2 ZP-OFDM System

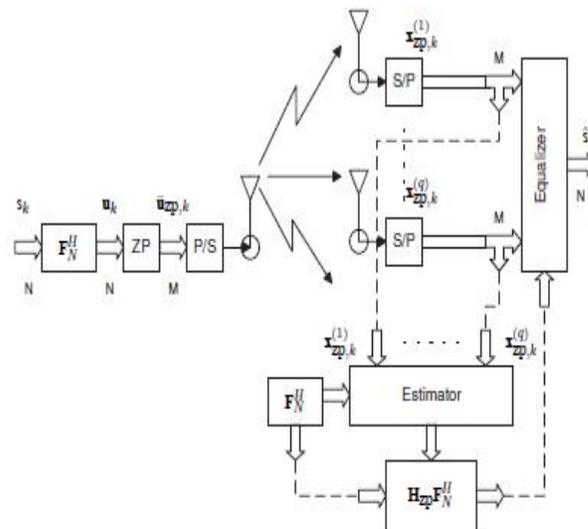


Fig. 4 - ZP-OFDM system: transmitter and receiver.

Figure 4 depicts the baseband discrete-time block equivalent model of a standard ZP-OFDM system. The only difference between ZP-OFDM and CP-OFDM is that the CP is replaced by L trailing zeros that are padded at each precoded block \mathbf{u}_k to yield $\mathbf{u}_{zp,k} = \mathbf{T}_{zp}\mathbf{u}_k = [u_k(0), u_k(1), \dots, u_k(N-1), 0, \dots, 0]^T$ where $\mathbf{T}_{zp} = [\mathbf{I}^T_N, \mathbf{0}^T_{L \times N}]^T$. We can write the received block symbol $\mathbf{x}_{zp,k}$ as

$$\begin{aligned} \mathbf{x}_{zp,k} &= \mathbf{H}_0 \bar{\mathbf{u}}_{zp,k} + \overbrace{\mathbf{H}_1 \bar{\mathbf{u}}_{zp,k-1}}^{\text{ISI}} + \mathbf{v}_{zp,k} \\ &= \mathbf{H}_0 \mathbf{T}_{zp} \mathbf{F}_N^H \mathbf{s}_k + \underbrace{\mathbf{H}_1 \mathbf{T}_{zp} \mathbf{F}_N^H \mathbf{s}_{k-1}}_{\text{ISI}} + \mathbf{v}_{zp,k} \end{aligned}$$

Where,

$$\mathbf{H}_0 = \begin{bmatrix} \mathbf{h}(0) & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{h}(L) & \dots & \mathbf{h}(0) & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{h}(L) & \dots & \mathbf{h}(0) & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{h}(L) & \dots & \mathbf{h}(0) \end{bmatrix}$$

$$\mathbf{H}_1 = \begin{bmatrix} \mathbf{0} & \dots & \mathbf{h}(L) & \dots & \mathbf{h}(1) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbf{h}(L) \\ \mathbf{0} & \dots & \dots & \dots & \mathbf{0} \\ \vdots & \ddots & \dots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & \dots & \mathbf{0} \end{bmatrix}$$

$$\begin{aligned} \mathbf{x}_{zp,k} &= [\mathbf{x}_{zp,k}^T(0), \mathbf{x}_{zp,k}^T(1), \dots, \mathbf{x}_{zp,k}^T(M-1)]^T \\ \mathbf{v}_{zp,k} &= [\mathbf{v}_{zp,k}^T(0), \mathbf{v}_{zp,k}^T(1), \dots, \mathbf{v}_{zp,k}^T(M-1)]^T \end{aligned}$$

and the key advantage of ZP-OFDM lies in the all-zero $L \times N$ matrix $\mathbf{0}$ which eliminates the ISI, since $\mathbf{H}_1 \mathbf{T}_{zp} \mathbf{F}_N^H = \mathbf{0}$. Forming the $qM \times N$ matrix \mathbf{H}_{zp} from the first N columns of matrix \mathbf{H}_0 , can be expressed as

$$\mathbf{x}_{zp,k} = \mathbf{H}_{zp} \mathbf{F}_N^H \mathbf{s}_k + \mathbf{v}_{zp,k}$$

Where,

$$\begin{aligned} \mathbf{x}_{zp,k} &= [\mathbf{x}_{zp,k}^T(0), \mathbf{x}_{zp,k}^T(1), \dots, \mathbf{x}_{zp,k}^T(M-1)]^T \\ \mathbf{v}_{zp,k} &= [\mathbf{v}_{zp,k}^T(0), \mathbf{v}_{zp,k}^T(1), \dots, \mathbf{v}_{zp,k}^T(M-1)]^T \end{aligned}$$

and \mathbf{H}_{zp} is, a block-Toeplitz matrix, defined as

$$\mathbf{H}_{zp} = \begin{bmatrix} \mathbf{h}(0) & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{h}(L) & \dots & \mathbf{h}(0) & \vdots \\ \mathbf{0} & \dots & \mathbf{h}(L) & \mathbf{h}(0) \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{h}(L) \end{bmatrix}$$

Corresponding to the first N columns of \mathbf{H}_{zp} , the \mathbf{H}_0 submatrix is block-Toeplitz and is always guaranteed to be invertible, which assures symbol recovery (perfect detectability in the absence of noise) regardless of the channel zero locations.

3 SS-BASED METHOD

The desire for a more efficient algorithm led to the development of subspace(SS) methods for the blind estimation of multi-channel FIR filters [12]. The basic idea behind these methods consists of estimating the unknown parameters by exploiting the orthogonality of subspaces of certain matrices obtained by arranging in a prescribed order the second order statistics of the observation. This scheme shares many similarities with well-known techniques for direction-of-arrival (DOA) estimation in a narrow-band array processing context. The existence of such SS-based methods for blind estimation was brought to light by Gurelli and Nikias [9] and Moulines et al [12] (see also Hua [11] and Abed-Meraim [1], [5],[6], [10]).

Blind channel estimation is particularly important for OFDM applications where severe ISI can arise from the time-varying multipath fading that commonly exists in a mobile communication environment. The varying channel characteristics must be identified and equalised in real time to maintain the correct flow of information. The use of SS-based methods to accomplish blind SIMO channel estimation for OFDM has been proposed for frequency-flat fading channels in [7], [8]. The extension of it to the general MIMO case has been successfully introduced by Zeng et al [15]. More recently, some SS-based methods have been proposed for single-user OFDM systems [13], [14]. The method in [14] can be applied to OFDM systems without CP and, therefore, leads to higher data-rate.

4 MNS-BASED METHOD

Unfortunately, a widely acknowledged problem with the aforementioned techniques is its extensive computational complexity due to the EVD of a 'large' dimensional matrix and rather slow convergence with respect to the number of block symbols. In fast changing environments, such as in cellular communications, their applications may be limited. This problem is alleviated by the MNS decomposition approach proposed by Abed-Meraim et al [4]. Based on this contribution, it is easy to show that, only $q - 1$ properly chosen noise eigenvectors are just as efficient as using the

whole noise subspace range(\mathbf{G}_{cp}) (or range(\mathbf{G}_{zp})) to yield a consistent estimate of \mathbf{H}_{cp} for the CPOFDM system (or \mathbf{H}_{zp} for ZP-OFDM system). Furthermore, each of the $q-1$ noise eigenvectors can be found by using EVD of a 'small' dimensional covariance matrix corresponding to the (distinct) pairs of channel outputs given by a *properly connected sequence* (PCS) defined as follows :

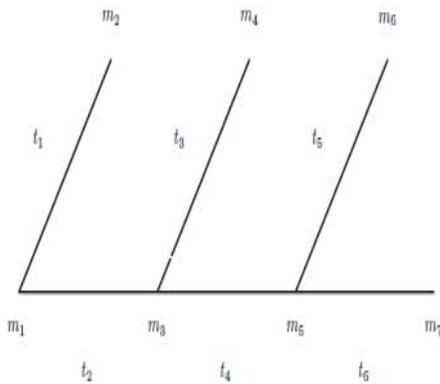


Fig. 5 - Tree that connects $q = 7$ channel outputs as its notes.

Definition 1: Denote the q system outputs by a set of members m_1, \dots, m_q . A combination of two ($q \geq 2$) members ($t_i = (m_{i1}, m_{i2})$) is called a pair. A sequence of $q-1$ pairs is said to be properly connected if each pair in the sequence consists of one member shared by its preceding pairs and another member not shared by its preceding pairs.

Example 1: Consider a system with one input and seven outputs. The following sequence of pairs has minimum redundancy (six pairs) and spans all system outputs m_1, \dots, m_7 .

$$t_1 = (m_1, m_2), t_2 = (m_1, m_3), t_3 = (m_3, m_4) \\ t_4 = (m_3, m_5), t_5 = (m_5, m_6), t_6 = (m_5, m_7)$$

Figure 5 demonstrates an example of PCS with $q = 7$. In the Tree pattern, the notes $m_2, m_4, m_6,$ and m_7 are ending nodes while the nodes $m_1, m_3,$ and m_5 are branching nodes.

Remarks:

- MNS-based method can be applied to applications relating to source localisation and array calibration [4].
- In practice a PCS is easy to construct, however, it is not a necessary condition to give the MNS.
- A PCS exploits the diversity of the system outputs with minimum redundancy. This follows, since a sequence has less than $q - 1$ pairs or a pair in the sequence has less than

two members, then the sequence does not give the required number of independent noise vectors.

- A set of $q - 1$ pairs span all the system outputs are not necessarily sufficient to give the required MNS.

5 SIMULATION RESULT

l	0	1	2	3	4
$h^{(1)}(l)$	-0.049+0.359i	0.482+0.569i	-0.558+0.587i	1.0000	-0.171+0.081i
$h^{(2)}(l)$	0.443+0.0384i	1.0000	0.321-0.194i	0.189+0.206i	-0.087-0.054i
$h^{(3)}(l)$	-0.211+0.322i	-0.198+0.918i	1.0000	-0.284+0.524i	0.138-0.190i
$h^{(4)}(l)$	0.417+0.093i	1.0000	0.873+0.145i	0.285+0.209i	-0.049+0.181i

Table 1: channel set :Impulse response

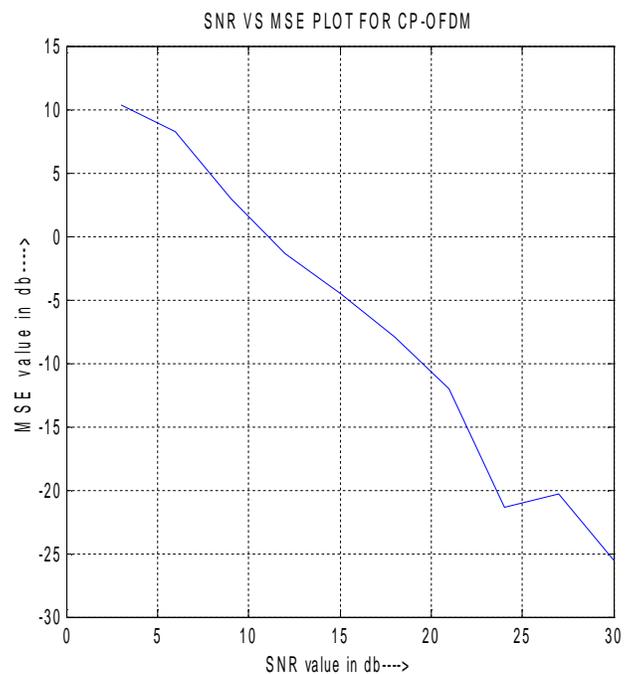


Fig .6 - Performance analysis of CPOFDM systems using SS-based method: SNR Vs MSE

Simulation Example 1: We first investigated the influence of weighting and performance of CP-OFDM receivers

through the implementation of the SS-based method in terms of their estimation capabilities. We fixed the number of OFDM symbols to $K = 1000$, and varied the SNR from 5 to 30 dB. $L=4$; $M=4$; $N=5$; $d=M+N$; L : antenna . M : channel length. N : smoothing. d : equalization delay

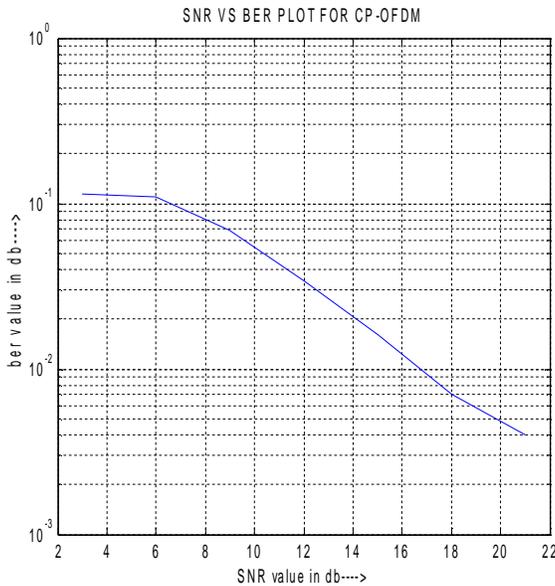


Fig .7 - Performance analysis of CPOFDM systems using SS-based method: SNR Vs BER

Simulation Example 2: The overall BER performance of the proposed SS-based method for the CP-OFDM systems Corresponding to SNR range of 2-20 dB. In order to check the equaliser gain.

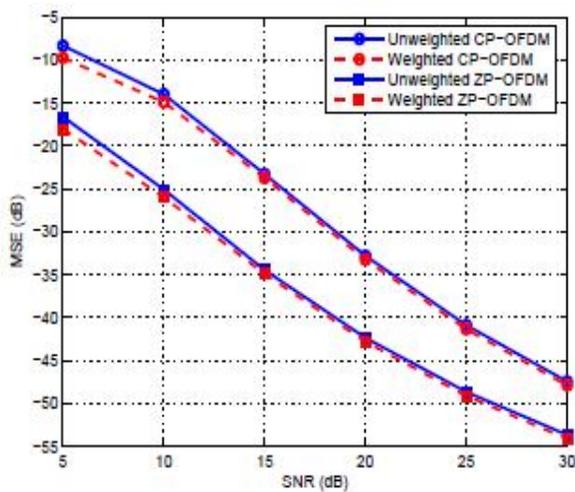


Fig .8 - Performance comparison of CP-OFDM and ZP-OFDM systems using the SS-based method: SNR vs MSE

Simulation Example 3: We first investigated the influence of weighting and compare the CP-OFDM and ZP-OFDM receivers through the implementation of the SS-based method in terms of their estimation capabilities. We fixed the number of OFDM symbols to $K = 1000$, and varied the SNR from 5 to 30 dB. We simulated the output of a SIMO with $q = 2$ FIR channels of maximum order $L = 4$. The generated symbols are transmitted through 20 sub-carriers and so the size of the FFT/IFFT was $N = 20$.

6 PROPERTIES OF THE PROPOSED TECHNIQUES

In all the aforementioned SOS-based methods for OFDM based SIMO systems, the focus has been on channel identification and equalisation. In this section we give some comments on the above proposed techniques. For uniformity, we subsequently drop the subscripts m/n and express the covariance matrices as $\mathbf{R}(i)$ (corresponding to SS-based method) and $\mathbf{R}(i)$ (corresponding to MNS / SMNS-based method). Moreover, we drop the subscripts cp/zp and express the block-Circulant matrix $\mathbf{H}_{cp,(i)}$ and block-Toeplitz matrix $\mathbf{H}_{zp,(i)}$ as $\mathbf{H}(i)$. For the sake of simplicity, we consider

$$\begin{aligned} q &= 2N \text{ CP-OFDM} \\ &= 2M \text{ ZP-OFDM} \end{aligned}$$

$$\begin{aligned} v &= qN \text{ CP-OFDM} \\ &= qM \text{ ZP-OFDM.} \end{aligned}$$

- The proposed channel estimators can be made to exploit the signal subspace regardless of the noise subspace and therefore the minimisation problem can be recast as a maximisation problem [5], [1], [12]. The solution of maximisation problem is considered more favorable to the minimisation problem, as there are fundamental limitations on the relative accuracy with which the smallest eigenvalues of the matrix can be computed, and they are more difficult to compute than the large ones. However, it is shown in [12] that the noise SS-based method exhibits better performance than the signal SS-based method.

- The main advantage of the MNS-based methods is that the large matrix EVD is avoided and the noise vectors are computed in parallel as the least eigenvectors of a smaller size covariance matrices $\mathbf{R}(i)$, $i = 1, \dots, q - 1$, which requires only $O(q^2)$ flops (in contrast with the $O(v^3)$ flops required for the computation of \mathbf{R}). Comparatively, the SMNS- and MNS-based methods have almost the same order of computational cost for SIMO systems .

- The proposed ZP-OFDM estimator requires the EVD of a data correlation matrix of size $m \times m$ to extract the orthogonal subspace. In contrast, the proposed CP-OFDM estimator requires the EVD of data correlation matrix of size $n \times n$. Since $m > n$, the proposed ZP-OFDM estimator is computationally more complex than the CPOFDM estimator.

- The CP-OFDM estimator is sensitive to channel zeros that are closed to the sub-carriers, whereas, the ZPOFDM

estimator guarantees symbol recovery and offers a superior BER performance.

- The proposed CP-OFDM based SIMO system rely on the usual insertion of CP as in standard OFDM systems. Therefore, it does not require transmitter modification and is applicable to all standardised OFDM systems. In contrast, the ZP-OFDM based SIMO system presented requires transmitter modification to introduce ZP redundancy by a filter-bank precoder. Note that ZP is used in the DVB standard in the form of guard bits.
- By using the optimal weighting matrix $\mathbf{W}(i)$, the identification procedure becomes quite insensitive to the ill conditioning problem. In fact, if a pair of channels have close common zeros, the corresponding block-channel Matrix $\mathbf{H}(i)$ becomes nearly singular and consequently $\mathbf{W}(i)$ becomes large. Therefore, the inverse of the weighting matrix, $\mathbf{W}(i)^{-1}$, will qualitatively provide 'more weighting' (i.e., larger weighting coefficients) to the noise eigen vectors associated with a well conditioned $\mathbf{H}(i)$ than to those corresponding to ill conditioned block channel matrix. However, weighting MNS (WMNS)-based method often incurs high complexity and involves large decoding delay, and does not trade well for the accuracy improvement [5].

7 CONCLUSION

This paper presents original reformulation of the SS-based estimation procedure for the blind identification of OFDM based SIMO FIR channels. It fully exploits the relations between the noise subspace of a certain covariance matrix formed from the observed signals. This reformulation provides some additional insights into the existing subspace algorithms. More importantly, it allows one to analyse the second order statistics of the output signals for the case of CP-OFDM and ZPOFDM receivers. This technique, although reliable and robust in some scenarios, require a computationally expensive and non-parallelisable EVD to extract the noise subspace. In fast changing environments, such as in cellular communications, their application may be limited and impossible (too costly) to implement. These problems are alleviated by a MNS based method which exploits a minimum number of noise eigenvectors for multi-channel identification. This technique of MNS, and especially the concept of PCS, turns out to be a powerful tool that can be applied to other OFDM-based SIMO systems. The proposed MN-based method is much more computationally efficient than the standard SS-based method at the price of a slight loss of estimation accuracy. However, better estimates of FIR channels can be obtained by a symmetric version of MNS with the same order of computational cost. Simulations have shown that the proposed method are effective and robust.

REFERENCES

- [1] K. Abed-Meraim, J. F. Cardoso, A. Y. Gorokhov, P. Loubaton and E. Moulines, "On subspace methods for blind identification of single-input multiple-output FIR systems," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 42-55, January 1997.
- [2] K. Abed-Meraim and Y. Hua, "Blind identification of multiple-input multiple-output system using minimum noise subspace," *IEEE Transactions on Signal Processing*, vol. 4, no. 1, pp. 254-258, January 1997.
- [3] K. Abed-Meraim and Y. Hua, "Weighted minimum noise subspace method for blind system identification," *ELSEVIER Digital Signal Processing*, vol. 69, pp. 15-27, May 1999.
- [4] K. Abed-Meraim, Y. Hua and A. Belouchrani, "Minimum noise subspace: concepts and applications," in *Proceedings of International Conference in Information, Communications and Signal Processing (ICICS)*, vol. 1, pp. 118-221, September 1997.
- [5] K. Abed-Meraim, P. Loubaton and E. Moulines, "A subspace algorithm for certain blind identifications problems," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 499-511, March 1997.
- [6] K. Abed-Meraim, P. Loubaton and E. Mouline, "Prediction error method for second-order blind identification," *IEEE Transactions on Signal Processing*, vol. 45, pp. 694-705, March 1997.
- [7] H. Ali, A. Doucet and Y. Hua, "Blind SOS subspace channel estimation and equalization techniques exploiting spatial diversity in OFDM systems," *ELSEVIER Digital Signal Processing*, vol. 14, pp. 171-202, 2004.
- [8] H. Ali, J. H. Manton and Y. Hua, "A SOS subspace method for blind channel identification and equalization in bandwidth efficient OFDM systems based on receive antenna diversity," in *Proceedings of the IEEE Workshop on Statistical Signal Processing (SSP)*, vol. 2, pp. 401-404, August 2001.
- [9] M. Gurelli and C. Nikias, "A new eigenvector based algorithm for multichannel blind deconvolution of input colored signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, pp. 448-451, April 1993.
- [10] H. Gazzah, P. A. Regalia, J. P. Delmas and K. Abed-Meraim, "A blind multi channel identification algorithm robust to order overestimation," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1449-1458, June 2002.
- [11] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Transactions on Signal Processing*, vol. 44, pp. 661-672, 1996.
- [12] E. Moulines, P. Duhamel, J. F. Cardoso and S. Mayrargue, "Subspace Methods for the Blind Identification of Multichannel FIR Filters," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516-525, February 1995.
- [13] B. Muquet, M. D. Couville and P. Duhamel, "Subspacebased semiblind channel estimation estimation for OFDM systems," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1699-1712, July 2002.
- [14] S. Roy and C. Y. Li, "A subspace blind channel estimation method for OFDM systems without cyclic prefix," *IEEE Transactions on Wireless Communications*, vol. 1, no. 12, pp. 572-579, July 2002.
- [15] Y. Zeng and T. S. Ng, "A semi-blind channel estimation method for multiuser multiantenna OFDM systems," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1419-1429, May. 2004.

Non-Restoring Divider Circuit Using a MCIT Based Adder Cell having Low Energy and High Speed Array

Owk Prasanth Kumar

Abstract--The paper discusses the design of 1-bit full adder circuit using MCIT. This proposed full adder circuit is used as one of the circuit component for implementation of Non- Restoring divider circuits. The proposed adder and divider schematics are designed by using SPICE and their layouts are generated. The divider circuits are designed by using standard NMOS and PMOS 180nm feature size and corresponding power supply 1.8 V. The parameters analyses are carried out by HSPICE analysis. We have compared the simulated results of the Shannon based divider circuit with CMOS adder cell based divider circuits. We have further compared the results with published results and observed that the proposed adder cell based divider circuit dissipates lower power, gives faster response.

Keywords-- CAS, CMOS, Logic gates, MCIT, Multiplexing, Propagation delay, Shannon theorem.

1. INTRODUCTION:

Digital dividers generally can be categorized as employing arithmetic operations to execute a division operation. Arithmetic dividers receive an input that combines the numerator and denominator. Look-up table implementations often require large look-up tables to be accurate for high-speed division, which is generally requiring significant processing time and chip space. Many look-up table implementations also require multiple iterations to improve accuracy, which increases latency associated with the division operation. The processing time required to perform the digital division operation, which corresponds to the amount of time required to perform the division process and referred as latency [1]. In many conventional digital divider designs, the latency determines the overall speed of the division process. As a result, most calculation type dividers typically provide only a few bits of precision for real-time operation.[1]

There are two kinds of array divider such as Non-Restoring Array Divider (NRAD) and Restoring Array Divider (RAD). This paper deals with Non-Restoring and Restoring array divider. The Non-Restoring array divider is guessing the quotient at each stage and when it is wrong it will not correct the remainder in this stage, instead of that it would continue to go to next stage. It has some extra remainder correction circuit after the last stage to correct the last remainder output by the divider. The Non-Restoring divider is much more efficient and faster than regular Restoring array divider. Since the array divider has many stages, it can be efficiently pipelined. More importantly, the Non-Restoring divider uses a very regular structure and each cell only needs to connect to the nearest neighbour cells, which makes it very efficient for VLSI design. The Restoring array divider circuit has controlled subtract cell which is used as the cell component of the divider circuit.[1]

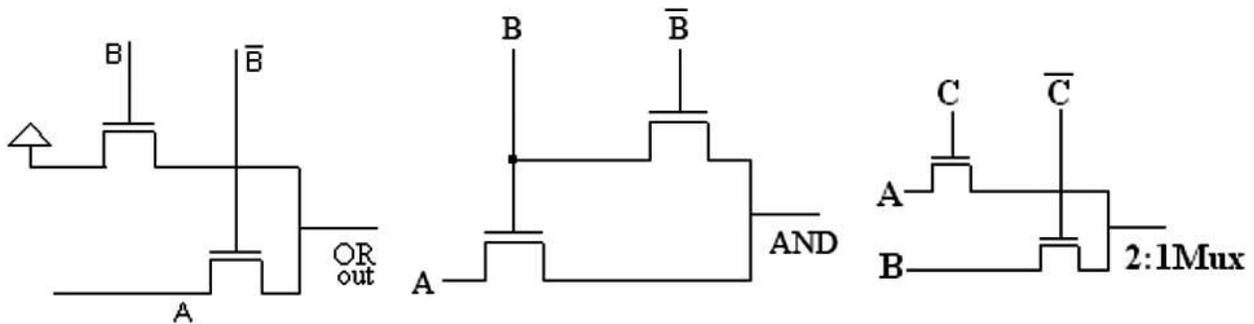


Fig.1:MCIT gates.

The Fig. (1.1) shows a basic AND, OR and 2:1 Mux circuit, that are designed using by Multiplexing Control Input Technique (MCIT). This logic gate design by MCIT

2. ADDER ARCHITECTURE FOR THE ARRAY DIVIDER:

The Shannon Theorem states that any logic expression can be expanded into two terms, the first with a particular variable setting a variable to 1, then multiplying it by the variable and then setting the variable to 0 and multiplying by the inverse. Shannon's Theorem can be stated in a generalized form as:

A function of many variables, $f(a_0, a_1, a_2, \dots, a_i, \dots, a_n)$ can be written as the sum of two terms, one with a particular variable (say a_i) set to 0, and other with it set to 1. $f(a_0, a_1, a_2, \dots, a_i, \dots, a_n) = a_i' f(a_0, a_1, a_2, \dots, 0, \dots, a_n) + a_i f(a_0, a_1, a_2, \dots, 1, \dots, a_n)$ [2]

2.1 ADDER CELL:

The proposed full adder circuit is designed by using Shannon theorem. The full adder sum and carry circuits are designed based on standard adder circuit

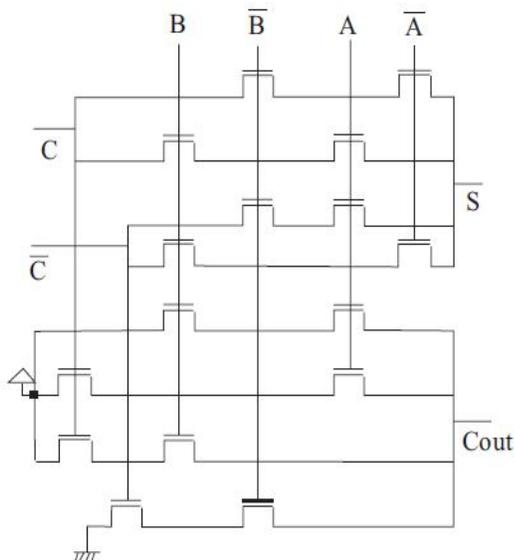


Fig:1.2 CAS cell.

equations. An input B and its complement are used as the control signal of the sum circuit which is shown in Fig. (1.2). The two-input XOR gate is developed using the multiplexer method. The output node of the two-input multiplexer circuit is the differential node. According to standard full adder equation, the sum circuit needs three inputs. In order to avoid increasing the number of

technique is reduced number of transistor than conventional CMOS circuit. The elaboration of this technique is gives the proposed Shannon full adder cell.

transistors due to the addition of a third input, the following arrangement is made; the CPL XOR gate is multiply with C's complement input and EXNOR gate is multiplied with input C and this reduces the number of transistors in the sum circuit. Compared with our previous paper, this kind of arrangement cause an increase in the number of transistors but this arrangement avoids the critical path delay. The C and C output node is called the differential node of the circuit. The differential node output is a summing output, as given in Standard full adder equation.[1]

The full adder carry circuit is designed by using fundamental Shannon equation. The source inputs are connected with logic '1', which results always 'ON' condition for the transistor. The actual inputs AB, BC and CA are connected in parallel to give the output $C=AB+BC+CA$. The circuit works according to the standard carry equation. In this circuit, all of the pass inputs are connected at VDD line so that the pass gates are always on. The control input terminals are connected the function inputs.

2.2 SUBTRACTOR CELL:

To find the value $(A-B)$, if assume that the input A is greater than B, we can use only 2's complement method, which is clearly shown in Fig. (1). In this method, 2's complement of B is to be added to A and the end around carry is to be ignored. All the bits of B is complemented to get 1's complement of B and then to the least significant bit B0 is added with '1' to get 2's complement of B. the right most full adder adds A_0, B_0 and '1'. The immediate left full adder circuit adds A_1, B_1 , and C_0 and so on. The end carry is ignored. The results $A-B$ is given by $DN\dots D1D_0$. Signed 2's complement is a modification of the sign-magnitude form in which addition and subtraction of the full adder circuit performed. The high order bit is still the sign bit and logical '1' still indicates a negative number[3]

3. NON-RESTORING DIVISION:

Binary division is basically a procedure to determine how many times the divisor D divides the

dividend A thus resulting in the quotient Q. At each step in the process the divisor D either divides A into a group of bits or it does not. The divisor divides a group of bits when the divisor has a value less than or equal to the value of those bits. Therefore, the quotient is either 1 or 0. A basic Non-restoring divider cell consists of a full adder and a XOR gate.[2]

It will read the practical remainder from the previous stage and depending on the quotient of the last stage it will add the divisor to obtain the remainder for next stage. In this paper the gate level schematic of full adder cells are used to obtain the FET level schematic, which pertains our proposed adder cells. After the design of one cell we can tile them in two dimensions to form a matrix. The Non-restoring array divider cell consists of full adder, and 2 inputs X-OR cells. The divider circuit can be roughly divided into two halves.[3]

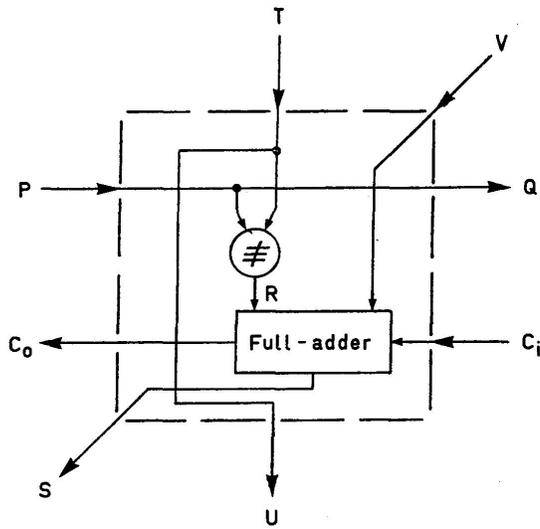


Fig.2: Controlled Adder Subtractor cell (CAS)

Figure 2 shows a single controlled adder-subtractor cell having four inputs T,V,C_i,P and four outputs S,C_o,U,Q. The cell function may be summarized as follows: for P = 0, E is the difference and F the borrow generated in forming B-(C+D), while for A = 1, E is the sum and F the carry produced by the addition B+C+D.[2]

4. NON-RESTORING ARRAY DIVIDER:

The design of an array is now described in which subtractions are performed by addition of the divisor in 2's complement form. In the basic cell, shown Fig. 4, a divisor digit presented at the input T is subjected to a controlled complementation, dependent on the state of the control. The worst case scenario for the delay was determined to be the one in which all inputs was set to logic high. The speed of the divider circuit can be determined by the

input P, before application to the input R of the full-adder. For P = 0, R = T, while for P = 1, R = T. The full-adder performs the addition V + R + Q to give the sum and carry outputs S and C_o. Also, independently of all other inputs, Q = P and U = T. In the Example, a denotes an addition and s a subtraction; sign digits are underlined for identification. The width of the array can be extended to accommodate input data of greater word length, and the number of rows in the array can be increased to give the quotient to as many places as may be required. The operating speed x of the array is given by [2]

$$T = Nt_d + N(N+1)t_b + t_n$$

Where t_d, t_b are the signal delays for the sum-difference output E and the carry-borrow output F of a single cell; t_n is the delay for a single negating gate. The dividend length, divisor length, quotient length are 2N, N, N+1 respectively.[2]

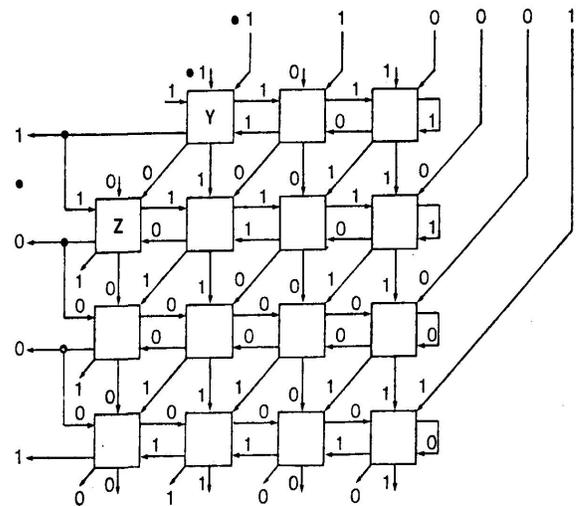


Fig 3 :6X4 divider using CAS cells of fig 2

5. ANALYSIS:

maximum delay along with critical path. A propagation delay of divider circuit for conserving power on a semiconductor chip is provided. The circuit includes a

delay chain responsive to a input signal for generating an output signal having a selectively adjustable delay at an output circuit. The delay chain techniques have been developed to reduce the energy dissipation of CMOS design systems. The minimization of power can be carried out by reducing the supply voltage, the capacitance, the number of transitions (e.g. the activity in the circuit) and by optimizing the timing of the signals. A large impact on energy is made by the supply voltage.[1]

By reducing VDD the energy dissipation decreases quadratically but the delay increases and the performance is degraded. A possible solution is that of using different supply voltages the higher voltage. Another technique is to compensate in different parts of the circuit. The parts not in the critical path are supplied by lower voltages, while the critical one by the loss of performance by replicating the hardware (parallelism) to keep the throughput.

5.1 HSPICE CODE:

****CAS****

```
.subckt cas a2 a21 b b1 c c1 e e1 18 20 15 16
Vd d 0 dc 1.8
```

*mux

```
M17 11 e b1 0 nmos l=180nm w=180nm
M18 11 e1 b 0 nmos l=180nm w=180nm
M19 d 14 12 d pmos l=180nm w=180nm
M20 12 14 0 0 nmos l=180nm w=180nm
M21 d 11 13 d pmos l=180nm w=180nm
M22 13 11 0 0 nmos l=180nm w=180nm
M23 d 13 14 d pmos l=180nm w=180nm
M24 14 13 0 0 nmos l=180nm w=180nm
```

*adder

```
M1 3 a21 c 0 nmos l=180nm w=180nm
M2 1 12 3 0 nmos l=180nm w=180nm
M3 4 a2 c 0 nmos l=180nm w=180nm
M4 1 14 4 0 nmos l=180nm w=180nm
M5 5 a21 c1 0 nmos l=180nm w=180nm
M6 1 14 5 0 nmos l=180nm w=180nm
M7 6 a2 c1 0 nmos l=180nm w=180nm
M8 1 12 6 0 nmos l=180nm w=180nm
M9 7 a2 d 0 nmos l=180nm w=180nm
M10 2 14 7 0 nmos l=180nm w=180nm
M11 8 c d 0 nmos l=180nm w=180nm
M12 2 14 8 0 nmos l=180nm w=180nm
M13 9 c d 0 nmos l=180nm w=180nm
M14 2 a2 9 0 nmos l=180nm w=180nm
```

*output nots

```
M25 d 1 15 d pmos l=180nm w=180nm
M26 15 1 0 0 nmos l=180nm w=180nm
M27 d 2 16 d pmos l=180nm w=180nm
M28 16 2 0 0 nmos l=180nm w=180nm
M29 d 1 17 d pmos l=180nm w=180nm
M30 17 1 0 0 nmos l=180nm w=180nm
M31 d 17 18 d pmos l=180nm w=180nm
M32 18 17 0 0 nmos l=180nm w=180nm
```

```
M33 d 2 19 d pmos l=180nm w=180nm
M34 19 2 0 0 nmos l=180nm w=180nm
M35 d 19 20 d pmos l=180nm w=180nm
M36 20 19 0 0 nmos l=180nm w=180nm
```

.ends cas

```
Va0 a0 0 0
Va01 a01 0 1.8
Va1 a1 0 0
Va11 a11 0 1.8
Va2 a2 0 0
Va21 a21 0 1.8
Va3 a3 0 0
Va31 a31 1.8
Va4 a4 0 1.8
Va41 a41 0 0
Va5 a5 0 0
Va51 a51 0 1.8
Vb b 0 0
Vb1 b1 0 1.8
Vb11 b11 0 1.8
Vb111 b111 0 0
Vb0 b0 0 0
Vb01 b01 1.8
Ve e 0 1.8
Ve1 e1 0 0
```

```
X1 a2 a21 b b1 e e1 e e1 18 20 15 16 cas
X2 a1 a11 b11 b111 20 16 e e1 l k l1 k1 cas
X3 a0 a01 b0 b01 k k1 e e1 m q0 m1 q01 cas
X4 a3 a31 b b1 q0 q01 q0 q01 o p o1 p1 cas
X5 18 15 b11 b111 p p1 q0 q01 r s r1 s1 cas
X6 l11 b0 b01 s s1 q0 q01 t u t1 u1 cas
X7 m m1 e1 e u u1 q0 q01 v q1 v1 q11 cas
X8 a4 a41 b b1 q1 q11 q1 q11 w x w1 x1 cas
X9 o o1 b11 b111 x x1 q1 q11 y z y1 z1 cas
X10 r r1 b0 b01 z z1 q1 q11 f g f1 g1 cas
X11 t t1 e1 e g g1 q1 q11 h q2 h1 q21 cas
```

```
X12 a5 a51 b b1 q2 q21 q2 q21 i j i1 j1 cas  
X13 w w1 b11 b111 j j1 q2 q21 s13 c13 s131 c131 cas  
X14 y y1 b0 b01 c13 c131 q2 q21 s14 c14 s141 c141 cas  
X15 f f1 e1 e c14 c141 q2 q21 s15 q3 s151 q31 cas
```

```
.tran 1 200 10
```

***180nm CMOS model files included from
<http://ptm.asu.edu/>

```
.end
```

5.2 CODE ANALYSIS:

The code is written for the Fig:1 for each CAS cell and calling it for the other CAS cells in the matrix of Fig:3. The main advantage of this code is the use of PMOS and NMOS and adding CMOS buffers near the output, thereby getting the outputs similar and accurate to the outputs of the general CMOS gates with a lesser number of gates thereby decreasing the propagation delay of the cell and hence the circuit. The Vcc is given a rudimentary voltage of 1.8 volts for simulation purposes.

The numbers from 1 to 36 represent the different nodes, the code is divided into different parts like the MUX

represented in the fig 1.2 the output nodes being the buffers at the outputs of the divisor and the remainder for the 6X4 matrix shown in fig.3. Further the adder subpart of the code represents the basic CAS cell which is used for the other 23 cells. The transient analysis of the output is observed and verified for all the 6 bit dividend and 4 bit non-zero divisor. The model files should be included in the code at the location shown. This code can be made to work for any CMOS technology by including the required model files. Minimum amount of glitches are observed during the transitions.

Table I. Power Dissipation, Propagation Delay, Area Throughput, Latency, EPI and Number of Transistor of the and, or and 2:1 Mux Circuits

gates	Parameter	0.5V [50nm]		0.7V [70nm]		1.0V [90nm]		1.2V [120nm]		2V [180nm]		2.5V [0.25 μ m]		3.5V [0.35 μ m]	
		Our	Conv	Our	Conv	Our	Conv	Our	Conv	Our	Conv	Our	Conv	Our	Conv
AND	PD (μ W)	0.014	0.062	0.049	0.072	0.110	0.318	0.386	0.462	0.755	0.645	3.85	1.238	4.157	7.04
	Delay (ps)	7	120	9	130	12	159	18	174	20	180	26	200	47	208
	Area (μ m) ²	28	70	32	96	50	144	66	180	220	462	350	729	880	1680
OR	PD (μ W)	0.044	0.072	0.055	0.169	0.416	1.242	0.942	1.57	0.997	2.662	1.799	4.239	6.6	8.6
	Delay (ps)	8	165	12	124	18	160	20	218	24	130	30	236	54	419
	Area (μ m) ²	28	80	32	96	50	144	66	180	220	462	350	729	880	1680
2:1 MUX	PD (μ W)	0.021	0.091	0.038	0.118	0.165	0.456	0.293	0.786	0.782	1.342	1.293	1.675	1.586	3.564
	Delay (ps)	5	70	6	96	8	104	14	110	16	116	22	132	26	148
	Area (μ m) ²	28	40	32	66	50	90	66	126	220	300	350	450	880	1680

(Our=proposed MCIT circuit Conv = conventional circuit (CMOS))

Table 2. Comparison of Power Dissipation, Propagation Delay and Area for 1-bit Adder Cell Designed with Different Technique

Adder Type	Supply Voltage	0.5V [50nm]	0.7V [70nm]	1.0V [90nm]	1.2V [120nm]	2V [180nm]	2.5V [0.25 μm]	3.5V [0.35μm]	5V [0.6μm]
Shannon	Power μW	0.042	0.371	0.442	0.569	1.984	12.04	24.88	58.6
	Delay (ps)	2	4	6	13	23	29	54	114
	Area (μm ²)	15x7	22x10	23x11	28x14	37x26	62x32	84x52	142x80
Mixed Shannon [13]	Power μW	0.309	0.437	0.513	0.658	2.529	14.8	31.13	84.05
	Delay (ps)	6	12	18	22	25	49	70	105
	Area (μm ²)	16x7	16x9	17x10	20x9	19x12	52x30	82x48	124x74
CPL	Power μW	0.817	2.247	3.249	4.907	9.936	14.782	29.69	48.75
	Delay (ps)	14	23	31	81	87	205	324	640
	Area (μm ²)	43x8	19x10	20x12	24x14	49x28	62x35	98x56	148x86
CMOS	Power μW	1.49	3.76	7.06	10.69	14.04	25.34	29.10	32.25
	Delay (ps)	654	854	971	995	1000	1401	1602	1995
	Area (μm ²)	25x7	53x10	67x14	74x11	93x26	117x32	186x52	280x80

Table 6. Non-Restoring Array Divider and Restoring Array Divider of Feature Size 0.35V: Comparison in Terms of Power Dissipation, Propagation Delay, Area, EPI, Latency and Throughput

Feature Size	Type of Adder	Power (mW)	Delay (ps)	Area μm^2	EPI (Watts/IPS)	Latency n	Throughput GHz
Non-Restoring Array Divider	Shannon	0.617	260	1512x76	0.794×10^{-9}	2.308	3.623
	Mixed Shannon	1.231	295	1562x89	0.7604×10^{-9}	8.72	1.8848
	CPL	1.977	2264	1818x105	0.788×10^{-9}	20.112	0.3977
	CMOS	2.7930	522	1858x141	0.7028×10^{-9}	16.176	1.259
Restoring Array Divider	Shannon	0.5773	296	2012x125	0.7788×10^{-9}	12.736	1.2562
	Mixed Shannon	0.5933	446	1542x93	0.7626×10^{-9}	18.624	0.859106
	CPL	0.3108	3226	1798x99	0.6981×10^{-9}	59.616	0.268384
	CMOS	0.669	664	2010x147	0.7154×10^{-9}	21.776	0.7347

**The table sources are from Low Energy, Low Latency and High Speed Array Divider Circuit Using a Shannon Theorem Based Adder Cell by Chinnaiyan Senthilpari, Krishnamoorthy Diwakar and Ajay K. Singh.

6. CONCLUSION:

Our proposed MCIT adder cell consists of 16 transistors and mixed Shannon based adder cell consists of 12 transistors compare to CPL-18 transistor and CMOS -28 transistors. The Non Restoring 6x4 bit divider circuits are simulated by using HSPICE TOOL. All the outputs checked from $1111111 \div 0001$ to $111111 \div 1111$. Various parameters such as area, propagation delay, dissipated power, throughput, latency and EPI are determined from array dividers layout of feature size 180nm. The divider circuits are analyzed using BSIM 4 parameter analyzer and results are given in Table 6. Our proposed Shannon adder based Non-Restoring array divider circuit gives less power dissipation, lower delay, low EPI, low latency and high throughput compared with mixed Shannon, CPL and CMOS based array divider circuits due to lower critical path in our proposed adder cell[1]. Similarly, dominance reflects in the restoring array divider circuit in terms of power, delay, EPI, latency and throughput. Our MCIT based Non-restoring array divider circuit shows an improvement of 12.16% in propagation delay, 2% in EPI,

81.87% in latency and 65.32% in throughput compared to Shannon based Restoring array divider circuit. [1]The Non-restoring divider circuit implemented by using our proposed MCIT based adder cell dissipates lower power than CPL, CMOS and mixed Shannon based adder cell divider circuits irrespective of power supply . [1]

The design has been presented for the non-restoring division in fig.3. Equations for the operating speeds of this array indicate that the logical design of the cells should be primarily achieve the rapid propagation of signals along the rows.

The following are the advantages:

1. Quotient digits are obtained from cell outputs without the use of extraneous gates.
2. The use of carry signal for control, rather than as a sum signal enhances the operating speed.
3. Use of CMOS buffers at only the outputs and PMOS and NMOS everywhere else reduces the number of logic gates required.
4. The fan out of the circuit drastically increase compared to the CPL circuits.

5. The array can be implemented using small number of transistors.

6. Finally as the number of gates decreases, even the propagation delay decreases hence increasing the speed of computation.

7. REFERENCES:

1. "Low Energy, Low Latency and High Speed Array Divider Circuit Using a Shannon Theorem Based Adder Cell"
Chinnaiyan Senthilpari, Krishnamoorthy Diwakar and Ajay K. Singh- Recent Patents In Nano Technology, 2009 .
2. H.H.Guild-"Some cellular logic arrays for non-restoring division".(The Radio and Electronic Engineer, Vol 39,1970.)
3. "An augmented iterative array for high speed binary division"
Marus Cappa and V.Carl Hamacher
IEEE Transactions of computers, Vol- c22,no:2, ,1973

Owk Prasanth Kumar is from Visakhapatnam, Andhra Pradesh and pursuing his 3rd Year of Bachelor degree(2009-13) in Electronics and Communication Engineering from Visvesvaraya National Institute of Technology(VNIT), Nagpur, Maharashtra ,India.

owkprasanthkumar@ece.vnit.ac.in
prashanth.bobby183@gmail.com

Stimuli Sensitive Interpenetrating Network

Pham Trong Binh and Hsu, Hao-Chun and Zhao Zhenyu

Abstract- Hydrogels are three-dimensional, cross-linked polymer networks that can absorb a large amount of water or solvent without dissolution. Due to their high absorbency, soft consistency and other properties, such materials can be used in medical industries to produce artificial organs, contact lenses, dressing for wound treatment and drug transporting devices. In this research poly(N-Vinylcaprolactam) (PVCL) hydrogels were synthesized with different cross-linkers. It was found that cross-linkers affected the stability of the hydrogel and those with high hydrophilicity enhanced its swelling capacity. Our experiments showed that ethylene glycol dimethylacrylate (EGDMA) was the ideal cross-linker. Hydrogels of an interpenetrating polymer network (IPN) composed of the thermo-sensitive PVCL and pH-sensitive poly(N-acryloyl-N'-ethylpiperazine) (PAcrNEP) were prepared via sequential polymerization. They were characterized for their thermo and pH-responsive behaviour by time dependence, equilibrium and oscillatory swelling studies. The results showed that these IPN successfully exhibited a combined pH and temperature-sensitivity at a temperature range of 20-35°C and a pH range of 3-6. This is a milestone in the study of hydrogels because most of them focused mainly on gels sensitive to basic medium only. The network also exhibited superior swelling capacity compared to pure PVCL even at high pH. Oscillatory swelling study showed that the IPN was able to respond to pH pulses quickly and reversibly. Dye absorption studies showed that IPN is a potential purifying agent. It was also found that the IPN is a good substrate for growing nanoparticle. Further research may include permeation studies of the IPN to test its ability for drug delivery.

Index Terms- Dye Absorption, Hydrogels, Interpenetrating Network, Nanoparticles, Polymers, Stimuli Sensitive, Swelling Synthetic Chemistry

1 INTRODUCTION

Hydrogels are three-dimensional, cross-linked polymer networks that can absorb a large amount of water or solvent without dissolution. Depending on the type of polymer that constitute the matrix, hydrogels can be responsive to various external stimuli such as pH, temperature, ionic strength, light, magnetic field, etc¹. Those that experience changes in swelling in solutions are called stimuli-responsive materials. Due to their high absorbency, soft consistency and other properties, such materials can be used in medical industries to produce artificial organs, contact lenses, dressing for wound treatment and drug transporting devices². In addition, hydrogels are found to be a very suitable material for stabilizing metallic nanoparticles in aqueous solutions by maintaining the nanostructure of the particles. Silver nanoparticles are nontoxic and exhibit good antimicrobial property³, but they have poor surface binding property, and this can be solved by growing hydrogel-embedded silver nanoparticles.

Furthermore, the dye absorption property derived from hydrogels can present an alternative way of treating industrial waste⁴.

Recently, work has been done to prepare hydrogels that are both pH and temperature sensitive by copolymerizing two monomers to form an interpenetrating polymer network (IPN)^{5,6}. Since there is no chemical bond between the two component networks, each network can retain its own unique property. While substantial research has been done on IPNs that function in basic medium, not many looked into those that might function in acidic medium, even though many biological processes take place in acidic environment, such as digestion in the stomach.

Poly (N-acryloyl-N'-ethylpiperazine) (PAcrNEP) is an ionisable hydrophilic polymer. Cross-linked PAcrNEP is able to swell in water and its swelling behaviour is greatly pH-dependent due to the ionization/deionization of the amine groups. At a high pH, usually greater than 5.5, the amine groups are not ionized, which keeps the PAcrNEP in its collapsed state. At low pH, the amine groups are ionized and the charged amine groups repel each other, leading to the swelling of PAcrNEP⁷.

Poly (N-Vinylcaprolactam) (PVCL) is a temperature sensitive IPN. Cross-linked PVCL exhibits drastic swelling transition at its lower critical solution temperature (LCST) of 35°C. At temperatures lower than 35°C the gels is swollen

- Hsu, Hao-Chun is currently pursuing a high school diploma at NUS High School of Math and Science, Singapore, (+65) 68970582, patrickhsu0905@hotmail.com
- Pham Trong Binh is currently pursuing a high school diploma at NUS High School of Math and Science, Singapore, (+65) 65161709, ptrongbinh@gmail.com
- Zhao Zhenyu is currently pursuing a high school diploma at NUS High School of Math and Science, Singapore, (+65) 65161709, h0930054@nus.edu.sg

whereas at temperatures higher than 35°C, the gel dehydrates to the collapsed state due to the breakdown of the delicate hydrophilic /hydrophobic structure⁸.

2 AIMS

1. To synthesize pH- and temperature-sensitive interpenetrating networks based on PAcNEP and PVCL
2. To investigate the swelling behaviour of IPNs with respect to pH, temperature, and ionic strength.
3. To study permeation behaviours of these IPNs under various temperature and pH condition.
4. To study heavy metal absorption capability of these IPNs.
5. To study the ability of these IPNs to grow silver nanoparticles.
6. To study the dye absorption ability of IPNs.
7. To characterize the prepared hydrogel-silver nanocomposites (HSNCs).

3 METHODOLOGY

3.1 Preparation of gels

3.1.1 Preparation of PVCL hydrogels

Hydrogels are made up of 3 components: the monomer, 2g N-vinylcaprolactam (VCL), the crosslinker, ethylene glycol dimethylacrylate (EGDMA) and the initiator, 4 wt% Azobis 4-cyanovaleric acid (ACA). The gels differed in their cross linker concentrations, which were set at 6, 9 and 12 wt%. The mixture was dissolved in 3.5mL ethanol and bubbled in nitrogen gas. The solution was heated in an oven set at 70 °C until a gel was formed. The gel was cut into disks of diameter 1cm and thickness 0.5cm and left to dry in a vacuum drier set at 70 °C⁴.

3.1.2. Preparation of IPN

To synthesize the IPN, a piece of hydrogel was soaked with an equal mass of AcNEP and EGDMA, before adding the initiator (4wt% ACA) and dissolving it in 1mL ethanol. The concentration of EGDMA added was the same as the cross-linker concentration in the given gel. The mixture was heated in an oven set at 70°C overnight. The gel was washed with ethanol before being cut into disks of diameter 1cm and thickness

0.5cm and left to dry in a vacuum drier set at 70 °C⁴. Image 2 shows the completed product.

3.1.3 Preparation of HSNCs³

A piece of 2wt% non-IPN / 2wt% IPN gel was soaked in 10ml of 0.05M AgNO₃ solution and left to equilibrate for 24 hours. After that, 10 ml of 0.05M NaBH₄ was added and the solution was put in the freezer set at -20 °C. This extracted the silver ions from silver nitrate and embedded it into the gel.

3.2 Swelling Tests

To characterize the IPNs, swelling tests in de-ionized water at different temperatures and pH solution were first carried out. The swelling ratio was calculated using the following equation⁴.

$$SR = \frac{\text{Swollen weight} - \text{dry weight}}{\text{Dry weight}}$$

3.2.1. Water Swelling Test

IPN was soaked in water for 2 hours. Its weight was taken every 10 minutes. This was to investigate the swelling behaviour of the IPN with respect to time.

3.2.2. pH & Temperature Test⁶

For the pH test, the IPN was soaked in solutions of different pH values ranging from 2 to 10. For the temperature test, the IPN was soaked in solutions at different temperatures ranging from 20°C to 50°C and left for 24 hours before their weights were measured. This was to investigate the swelling behaviour of the IPN with respect to pH and temperature.

3.2.3 Ionic Strength Swelling Test⁶

IPN was soaked in NaCl solutions with different molar concentrations ranging from 0.005 to 0.500 M and left for 24 hours before their weights were measured.

3.2.4. Oscillatory Swelling Test

IPN was soaked in a pH 10 buffer solution for an hour, then transferred to a pH 3 buffer solution for the next hour, then finally put in the pH 10 solution for another hour. Weight of the IPN was measured every 10 minutes.

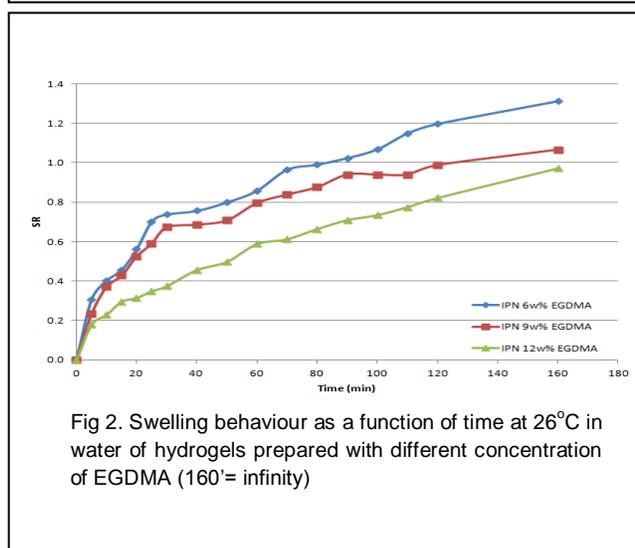
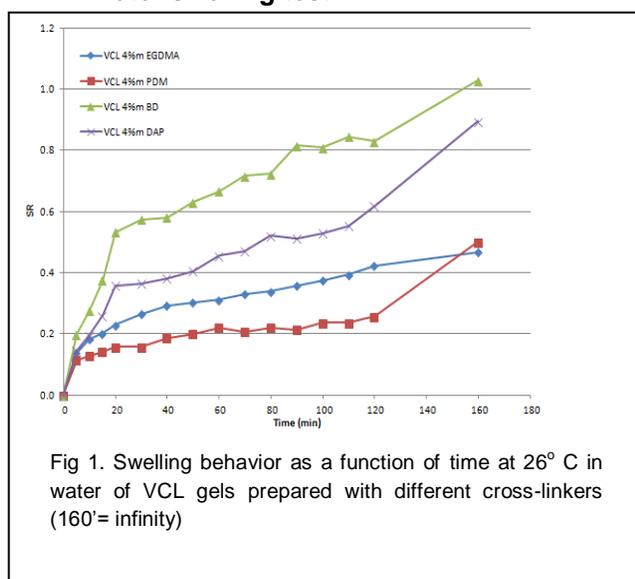
3.3 Dye Absorption Test

0.1 mM solutions of Congo red were prepared. The absorbance of the solutions was recorded on UV/Vis Spectrophotometer (Perking Elmer Lambda 25). A disk of IPN/PVCL of 1cm in diameter was soaked in each solution and left for 24 hours. The new absorbance was then recorded. The samples were labelled as followed: dye solution (D-P), dye solution with soaked PVCL (D-VCL), dye solution with soaked IPN (D-IPN).

4 RESULTS AND DISCUSSION

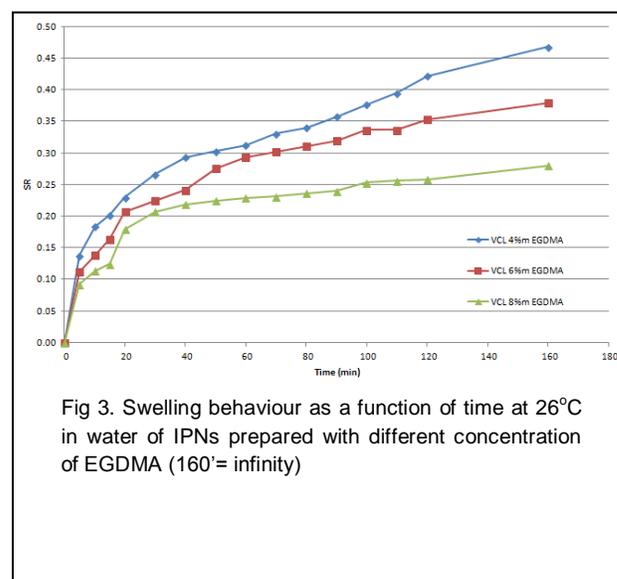
4.1. Swelling Test

4.1.1 Water swelling test



Hydrophilicity of the cross-linker played an important part in controlling the swelling of hydrogels. Figure 1 shows that an increase in hydrophilicity led to an increase in the hydrogel's swelling capacity. Even though gels prepared with BD and DAP as cross-linkers had a high swelling ratio, they tend to break apart easily during swelling. Therefore, EGDMA was chosen as the cross-linker in subsequent experiments because the gel prepared with EGDMA had a moderate swelling capacity and great stability.

It was found that the cross-linker concentration also had a significant impact on the swelling capacity of the hydrogel. Figure 2 shows that a higher cross-linker concentration led to a lower swelling capacity, because a higher amount of cross-linker rigidified the polymer network which then exhibited poor elastic response. Such a structure would hinder the diffusion of solvent into the network. The same can apply to IPN, as seen from the trend in figure 3.



Equilibrium swelling studies indicated that the IPN samples were capable of responding to external stimuli: pH, temperature and ionic strength.

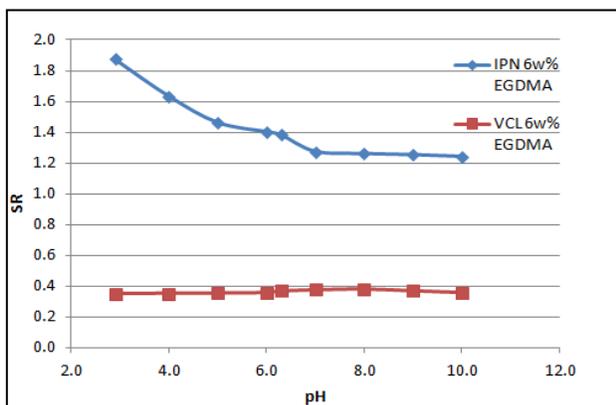


Fig 4. Equilibrium swelling behaviour as a function of pH of 6w% IPN and VCL hydrogels at 25°C.

Figure 4 shows the influence of pH on the swelling capacity of the IPN and pure VCL hydrogels from pH 3 to pH 10. The swelling curve of the IPN exhibited significant transition from pH 3 to pH 6: the swelling ratio decreased from 1.9 to 1.4. This transition was a result of the nature of the IPN network. In a buffer solution of low pH, the amino nitrogen in the AcrNEP component of the network was ionized, leading to the formation of like charges within the gel network. The repulsion of these charges resulted in the swelling of the IPN, which increased the water uptake capacity. The pKa of AcrNEP was 4.58; hence excess ionization was expected to occur around a pH of 3–4.5, which corresponded to the steep portion of the curve. In contrast, pure VCL gel exhibited little change in swelling capacity since they could not be ionized. It was also noted that IPN had better swelling capacity compared to PVCL gels even at high pH.

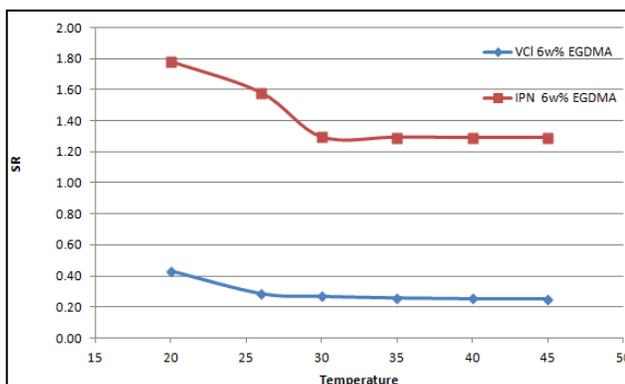


Fig 5. Equilibrium swelling behaviour as a function of temperature of 6w% IPN and VCL hydrogels in water.

Figure 5 shows the effect of temperature on the swelling behaviour of IPN and VCL hydrogels. Deswelling transition in the temperature range of 20°C to 35°C was observed for both samples and the transition curves were rather similar since the VCL component in the IPN was responsible for the thermo-sensitivity. However the transition of the VCL gel was sharper, with its swelling ratio almost halved from 20°C to 25°C. At a low temperature the gels were able to swell, leading to a high swelling ratio. As temperature increased, the delicate hydrophilic/hydrophobic balance in the VCL network was broken, leading to dehydration. In the IPN, the existence of AcrNEP network might have interfered with the hydrophilic/hydrophobic balance of the VCL network, so dehydration might not have taken place at the same time, resulting in a gradual deswelling transition.

Figure 6 shows the influence of ionic strength on swelling behaviour. As the concentration of NaCl increased, the swelling ratio dropped drastically. This was due to the presence of counterions at high ionic strength, which hindered ionization of the amino nitrogen.

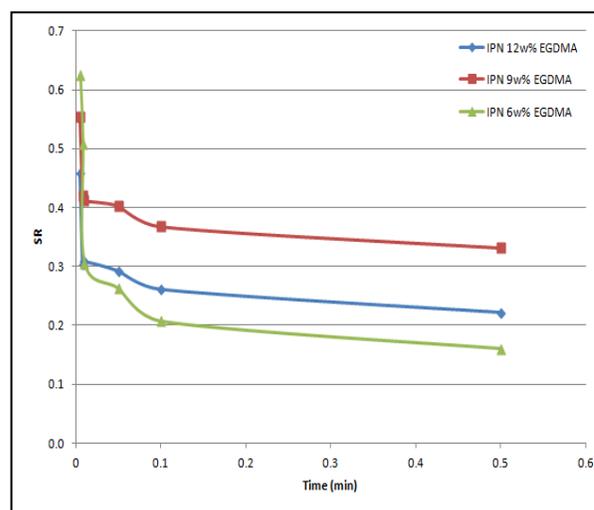
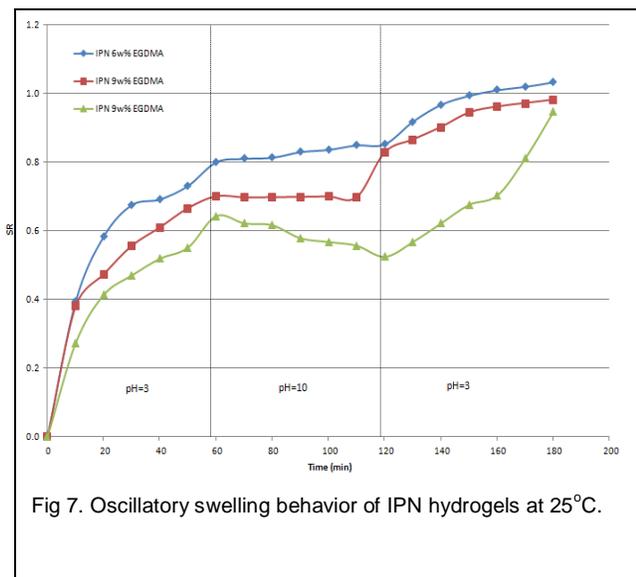


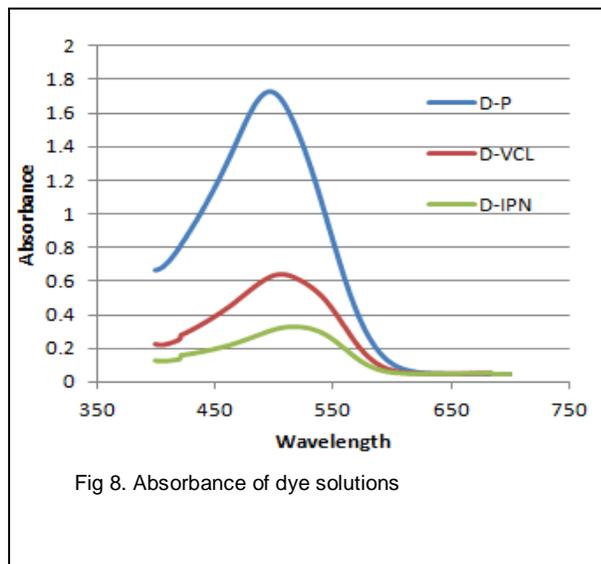
Fig 6. Equilibrium swelling behaviour of IPN and VCL hydrogels in water at 25°C.

4.1.2. Oscillatory swelling test



Oscillatory swelling study was carried out to investigate the ability of the IPN to respond to pH pulses. Fig 7 shows that the IPN gels were able to respond to changes in pH and the responses were quick and reversible.

4.2 Dye absorption test



Since Congo red dye exists as anions in solution, it is strongly attracted to the positive charged amine groups in the IPN. The change in absorbance (fig 8) verified the ability of the IPN to act as a potential purifying agent for waste contaminated with dye. The IPN was also found to be a superior agent compared to PVCL.

5 CONCLUSION

Temperature-sensitive PVCL and pH-sensitive PAcrNEP interpenetrating networks (IPN) were synthesized by sequential polymerization. The most suitable cross-linker used to synthesize the IPN was determined to be EGDMA, as it gives the IPN moderate swelling capacity and good stability. The IPN was shown to have superior swelling capacity compared to PVCL. It was also determined that high cross-linker concentration will lead to low swelling ratio of the gel. The IPN was shown to exhibit swelling transition at a temperature range of 20-35°C and pH range of 3-6. This indicated that each network is able to retain its properties in the IPN. Presence of ion salt was also found to hinder the IPN's swelling capacity.

6 APPENDIX



Image 1: Dry PVCL hydrogel



Image 2 Dry IPN



Image 3 Swollen IPN in pH 3

7 ACKNOWLEDGEMENTS

The authors express gratitude for guidance and equipment from Associate Professor Dr. Roshan Deen of National Institute of Education, Singapore.

8 REFERENCES

- [1] Xiaohu Xia, B.E., Fabrication and light scattering study of multi-responsive nanostructured hydrogels and water soluble polymers, University of North Texas, 2003
- [2] Zbigniew Ruszczak, Robert A Schwartz, Ewa Joss-Wichman, Surgical Dressings, Medscape Reference (2011)
- [3] Sangphil Park, P.S. Keshava Murthy, Saemi Park, Y. Murali Mohan, Won-Gun Koh, Preparation of silver nanoparticle-containing semi-interpenetrating network hydrogels composed of pluronic and
Preparation of silver nanoparticle-containing semi-interpenetrating network hydrogels composed of pluronic and pol (acrylamide) with antibacterial property, Journal of Industrial and Engineering Chemistry (2011) 293-297
- [4] S.J. Allen, B. Koumanova, Decolourisation of water/wastewater using adsorption (review), Journal of the University of Chemical Technology and Metallurgy 40-3 (2005) 175-192
- [5] Y. Murali Mohan, Kurt E. Geckeler, Polyampholytic hydrogels: Poly (N-isopropylacrylamide)-based stimuli responsive networks with Poly (ethyleneimine), Reactive and Functional Polymers 67 (2007) 144-155
- [6] Jing Zhang, Nicholas A. Peppas, Synthesis and Characterization of pH- and Temperature-Sensitive Poly (methacrylic)/Poly (N-isopropylacrylamide) Interpenetrating Polymeric Networks, Macromolecules 33 (2000) 102-107
- [7] G. Roshan Deen, Swelling Behavior and Metal-Ion Uptake Capacity of pH-Responsive Hydrogels of Poly (N-acryloyl-N'-ethylpiperazine), Journal of Dispersion Science and Technology 31 (2010) 1673-1678
- [8] Volodymyr B. Boyko, N-Vinylcaprolactam based bulk and microgels: synthesis, structural formation and characterization by dynamic light scattering, Doctoral dissertation, Dresden University of Technology (2004)

Implementation of AODV protocol with and without fuzzy logic for reliable multicast routing in adhoc networks

R.Senthil Kumaran

Abstract — A major aspect of adhoc networks is that the nodes can move randomly, which requires the routing protocols in adhoc network to quickly respond to the network topology change in order to guarantee successful data packet delivery. Multiple routing paths are established to provide extra schemes of video streaming or multicast transmission and enhance the robust transmission against unreliability and limited bandwidth of wireless links. In this paper, AODV routing protocol can be modified with fuzzy logic named fuzzy modified AODV routing (FMAR) protocol for multicast routing in mobile adhoc networks. The fuzzy logic weighted multi-criteria of the protocol is used to dynamically evaluate the active route life time in order to determine the appropriate routes. Due to frequent node movements, the topologies of mobile adhoc networks change rapidly. The Fuzzy rule base depends upon the number of hop counts, sent controlled packets and the energies of the nodes on the routes. The enhancement of FMAR protocol was implemented for quickly maintain and repair the routes with the dynamic Lifetime of the routing table before they crashed. The Fuzzy rule base depends upon the number of hop counts, sent controlled packets and the energies of the nodes on the routes. The simulation results of enhancement of FMAR protocol will be efficient than FMAR protocol with respects to the node mobility, the packet delivery ratio, average route acquisition latency delay, the routing overhead and the average end to end delay.

IndexTerms — AODV Protocol, FMAR protocol, Fuzzy logic, Multicast routing, Network lifetime, Network simulator

1 INTRODUCTION

An adhoc network is a collection of mobile nodes. Usually it has no centralized or fixed infra structure. Therefore its topology changes frequently. The main characteristics of adhoc networks are node mobility and node power control practices. The routing paths that consist of a number of wireless links are frequently blocked and then reestablished during time period of packet transmission. Hence the quality of wireless links often fluctuates due to the channel fading and the interference from other communications. Owing to the unexpected congestion or block of routes and the limited bandwidths of wireless links, it is hard to guarantee the packet delivery and it is also hard to provide good qualities of multicast. To overcome this defect, establishing multiple paths between a sender and receivers. AODV is an on-demand distance vector routing protocol. The protocol is well known for the use in adhoc networks. Our protocol in this paper is a fuzzy modified aodv for multicasting.

In aodv, the lifetime field of the routing table entry is static, which is either determined from the control packets or initialized to route lifetime (LT)[9]. AODV has a table-driven routing framework and a set of destination sequence numbers. It relies to a certain extent on timer-base activities

(routing table entries expiration timers). If a route is not used recently, the entry is expired. AODV has a mechanism to expired stale route or prefer “fresher” route when it faces with multiple choices[8]. For route discovery, the source broadcast a RREQ to find route to destination, the receivers follow the reverse paths pointing towards the source. The destination replies RREP back to the source; the reply packets including some information are generated by intermediated node following the precursor list. For route maintenance, an existing routing entry may be invalidated to the destination node. In that case, the invalidation is propagated to their next hop node. For route repair, when an intermediate node moves, its upstream neighbors will propagate RERR message to each of its active upstream neighbours.

Concerning the multicast, various protocols have been proposed for adhoc networks[3]. They can be classified in to two types according to their configurations. One of the two types is tree-based (e.g., Adhoc multicast routing (AMRoute), and Adhoc Multicast routing protocol utilizing increasing id-numbers (AMRIS) [10] and the other is mesh-based (e.g., on-demand Multicast routing protocol (ODMRP) and core-assisted mesh protocol(CAMP). There are many investigations in the area of multipathsR. transmission (MPT)[16][2]. Although the selection of an optimal path set is an NP-complete problem. The split multipath routing (SMR) protocol [2] as an on-demand routing scheme built maximally disjointed multi-paths, which efficiently utilizes the available network resources to simplify the route recovery process and minimizes control

-
- R.SenthilKumaran, M.Tech, Asst.Prof/ ECE, V.R.S college of Engineering and Technology, Villupuram, TamilNadu, India. E-mail:sen19841@gmail.com

message overhead. The SMR always constructs two maximally disjoint routes from a source to a destination after finishing the routing discovery. In a word, the use of static discovery to determinate multi-paths is hard to cope with instantaneous environmental variation in an adhoc network. Due to the complexity of the determinate variable route life time of the multipaths, only a few network researchers attempted to dynamically evaluating the route lifetime parameters in order to determinate route selection, maintain and repair.

Referring to the route lifetime, the route life time is equal to zero when the route is active during transmission, and it should be deleted at the end of the transmission, and that the route lifetime is equal to infinity when the route is inactive. In AODV, route lifetime (LT) is equal to a predetermined static value and in milliseconds. Route life time is the period of time that the route can stay active in the routing table. It is restricted and equal to an adaptive value[15] proposed that the path between two nodes should be remained in the routing tables along as the minimum parameter (route lifetime) is greater than a certain threshold. They stated that the route life time is an un-restricted adaptive lifetime, introduced the adaptive route lifetime method to minimize routing delay and overhead. They used mathematical tools to determine the values of adaptive route lifetime. The node movement and node power consumption are difficult to be predicted, the dynamical evaluation of the route lifetime is better than that statically assigned by AODV. The protocol applied a fuzzy logic system to dynamically evaluate the route expiry time. The fuzzy logic is chosen because there are uncertainties associated with node mobility and the estimation of link crash, moreover, there is a mathematical model can be capable of estimating the node mobility. In addition, this fuzzy multicast routing protocol is satisfactory to take some controlling factors (such as the node remained energy) in to consideration. Therefore, our protocol is a multi-criteria fuzzy evaluation for multicast routing protocol. In this paper, the packets are multicast via multiple paths to multiple receivers of a multicast group instead of successively performing single cast in adhoc networks. The technique determinates the two most stable routes from a source to a destination. One of the routes acts as the main path and the other path for extra function route. Thus, when the main path is unreliable, it can be replaced immediately by the preferred alternative. For constructing our multicast routing protocol, modify the approach of AODV to establish the two most stable paths [3].

The Fuzzy Logic Weighted Multi Criteria determines the appropriate routes to a multicast group. The constructed protocol is called FMAR (fuzzy modified AODV multiple routing). The advantages of FMAR are that a source can apply the multiple routes to successfully communicate with its receivers and that each multicast

route has mesh topology to deal with the suddenly crashed links during transmission. The FMAR make route decisions according to the replied messages containing the information of remained energies of the nodes on the routes, the number of hop-counts and the sending control packet of the intermediate node on the route. The FMAR applies a set of definitions (membership functions) and a set of rules (rule-bases) to select a most appropriate main route and an alternative route and let them stay active in the routing table. The route maintenance and repair will be handle in accordance with route life time information to keep multicast smooth going in adhoc network, so that the packet delivery may be reliable.

2 MULTICAST ROUTING PROTOCOL WITH MULTI-PATHS

In adhoc networks, the goal of establishing multiple routing paths is to solve the problem of the sudden breaks of routing paths from a source to receivers. The fuzzy modified AODV multiple routing (FMAR) is designed for the above purpose. The design procedure is as follows: Fig. 1 illustrates how the source S1 establishes a multicast group including two receivers (R1, R2). The forwarding nodes forward the requests and relay the reply. In the request phase, a source node (e.g. S1) floods a RREQ packet as an advertising packet throughout the entire network. The next nodes receive a non-duplicated RREQ and record the upstream node address (S1) in its routing table, and then re-flood the RREQ. Finally, after accepting the RREQ, the multicast receivers reply RREPs to the source S1. The receiver waits for a suitable amount of time so that all possible routes can be gained. In the reply phase, a receiver traces back by increasing Hop Counts along the reverse paths to the source to record the energy and the number of control packets of intermediate nodes. By the use of fuzzy logic weighted multi-criteria, the source selects the most stable route with maximum route lifetime instead of selecting the minimum delay time route. The remaining routes are stored in the routing table to be used as the second routes to establish the most stable route if it crashes.

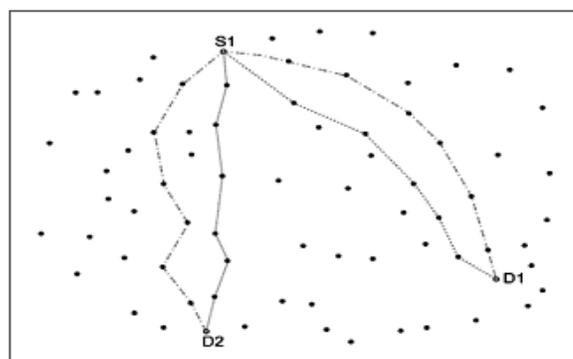


Fig. 1. The source S1 select two comparatively stable disjoint routes for destination D1 and D2, respectively.

3 FUZZY RULE BASE AND METHOD FOR DETERMINING ROUTES

In this section, we modify AODV and introduce the fuzzy logic weighted multi-criteria for assigning active route lifetime (LT) of each entry in the routing table. For the purpose of reducing the precious bandwidth and the overhead of transmission via void, the routes are determinate dynamically. Therein, a reliable route should be a route with maximum expiry time and should be substituted when it is unavailable. A source node in the ad hoc network has multi-paths to each destination. If the route is not used for a period of time or the lifetime is small, the route entry has high probability to expire due to high probability of its instability. The route will be partially repaired when the lifetime of the route is going to be less than a certain threshold. This approach is useful for computing route lifetime by the different criteria under the fuzzy environment. The practical problems are often characterized by several non- commensurable and competing (conflicting) criteria. The considered criteria can be based on the evaluation of the route characteristics including HopCount, the number of control packets and the lowest energy of the route node. Having defined the fuzzy linguistic rules, the HopCount is an evaluation criterion for Lifetime (active remain time in the routing table) and is described as the number rating of nodes along the route between the source and destination. When the HopCount is high, the probability of route broken is also high because of node's mobility. Therefore, the time that the path remains in routing table (the lifetime) should be smaller, thus the rating of Lifetime (expressed by LT) will be given small similarly. Consequently, the rules should be as follows: membership functions corresponding to each element in the linguistic set (HopCount, SntCtrlPkt, EngyMin and Lifetime) must be defined. We present the method to design its membership functions. The criteria rating can be assessed by linguistic terms (dimensionless index) such as very low (VL), low (L), medium (M), high (H) and very high (VH). The linguistic rating scale applied is illustrated in Fig. 2, and the membership functions of the five linguistic values are shown in Fig. 3.

- H1: If Hop Count is very high, then LT is very low.
- H2: If Hop Count is high, then LT is low.
- H3: If Hop Count is medium, then LT is medium.
- H4: If Hop Count is low, then LT is high.
- H5: If Hop Count is very low, then LT is very high.

The membership functions for Hop Count and LT rating value are expressed with dimensionless index within [0, 1] as shown in Fig. 2. In general, the LT is an inverse ratio to Hop Count. SntCtrlPkt is sum of the number of sent packets and the number of received packets. It is also a valid factor to evaluate the LT of the route entry in the routing table.

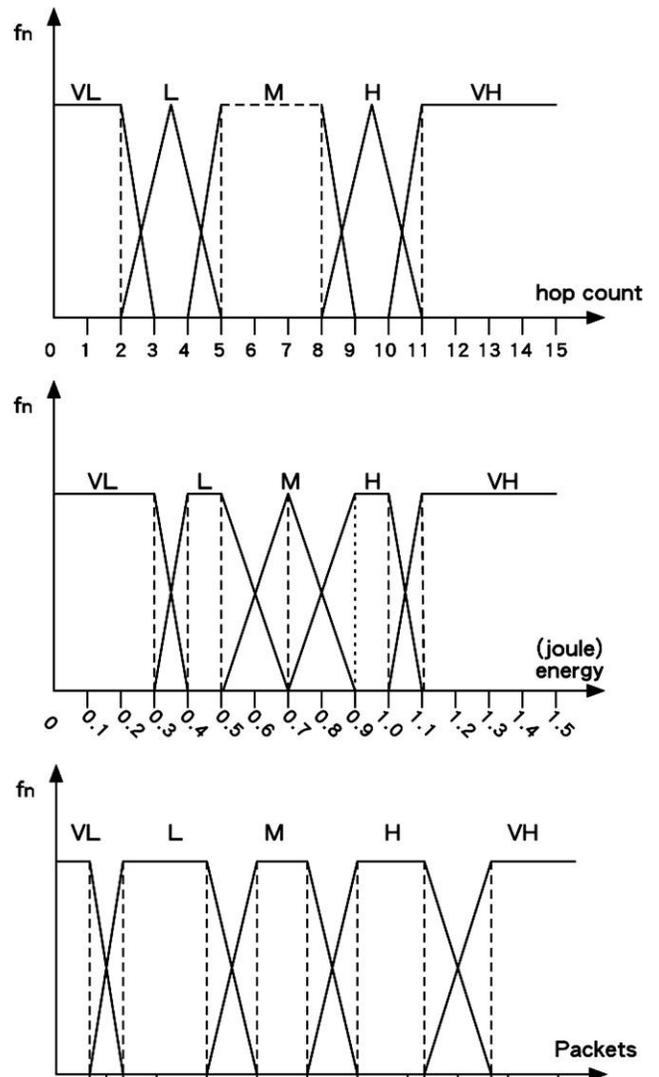


Fig. 2. Rating scale for accessing hop count, sent ctrlpkt and energymin criteria (VL, very low; L, low; M, medium; H high; Vh-very high)

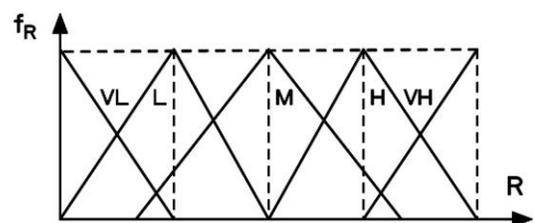


Fig. 3. Membership functions for linguistic rating values (VL, very low (0, 0, 0, 0.3); L, low (0,0.3,0.3,0.5); M, medium (0.2, 0.5, 0.5, 0.8); H, high (0.5, 0.7, 0.7, 1); VH, very high (0.7, 1, 1, 1))

If the transmission is interrupted or the nodes move frequently, then the link would very probably break, and

then the control packets would increase. When the route is congested, the shared bandwidth of channels would be reduced, so that the route became unstable due to the retransmission of the lost packets. The control packets can be any one of the following types: HELLO, RREQ, RREP, RERR and RREP_ACK. If the RREP packet records the number of sending control packet of the intermediate nodes is high, it represents the route is probably unstable. So, high number of these packets means high probability to lose some of its current links or packets.

The rules for the relationship between SntCtrlPkt and LT are:

- N1: If SntCtrlPkt is very high, then LT is very low.
- N2: If SntCtrlPkt is high, then LT is low.
- N3: If SntCtrlPkt is medium, then LT is medium.
- N4: If SntCtrlPkt is low, then LT is high.
- N5: If SntCtrlPkt is very low then LT is very high

Similarly, The SntCtrlPkt rating value is expressed with dimensionless index within [0, 1] and the LT is an inverse ratio to SntCtrlPkt. In addition, the energy characteristic can also be considered as an evaluation criterion. We use the energy consumption rate as the parameter to describe the power condition of an arbitrary node i. We denote the power consumption rate as PC_i. The value of PC_i is evaluated with a very simple method that is similar to that in[2]. Since the power consumption of a node is caused by the transmission, reception, and overhearing of packet activities, we define PC_i by the equation as below:

$$PC_i = \frac{(W_r \times M_r + W_s \times M_s + W_o \times M_o)}{T}$$

$$RT_i = \frac{RP_i}{PC_i}$$

Where PC_i = Power consumption rate.

W_r = Power consumed by the network interface when node i receives a packet

W_s = Power consumed by the network interface when node i sends a packet.

W_o = Power consumed by the network interface when node i overhears a packet.

T = Time period during the node i consumes its energy.

M_r, M_s, M_o are the amount of received, sent and overhears packets.

RP_i = Remaining battery energy of the node i.

RT_i = Residual time of the node i.

The life time of the route is related to, Min RT_i = Energy min.

If the EngyMin is low, the probability of the link broken will be high. Thus, the rules for the relationship between EngyMin should be as follows:

- E1: If EngyMin is very high, then LT is very high
- E2: If EngyMin is high, then LT is high
- E3: If EngyMin is medium, then LT is medium
- E4: If EngyMin is low, then LT is low
- E5: If EngyMin is very low, then LT is very low

Contrarily to the relationships between Hop Count

and LT and between SntCtrlPkt and LT, the LT is a direct ratio to the EngyMin. Considering the former two criteria, the membership function for EngyMin criterion should be considered as the dimensionless index [0, 1]. Based on this concept, the reverse energy limit, RvEngyMin, is defined as RvEngyMin = (1-EngyMin).

The membership functions for the rating values of EngyMin and RvEngyMin are respectively expressed by the two equations as below:

The value of EngyMin = (a, b, c, d) and

The value of Rv EngyMin = (1-d, 1-c, 1-b, 1-a).

This rule-base is to be employed to build the fuzzy system as described below:

The LT ratings of the different routes to the same destination under weighted multiple criteria are evaluated. The membership function of rating values is assigned as F(i) for the route i. Here, we respectively correspond the membership function H(i), S(i), E(i) and RvE(i) to Hop Count, Sent-CtrlPkt, EngyMin and RvEngyMin. The membership functions E(i) and RvE(i) of the EngyMin rating value are expressed as follows:

E(i)=(a,b,c,d) and RvE(i)=(1-d,1-c,1-b,1-a)

The above rules can be combined with a rule base as represented in below table:

Table 1

The ratings of the alternative routes according to the three criteria described above

	HopCount	SentCtrlPkt	EngyMin
R1	VH	H	L
R2	M	VL	M
R3	H	L	VH
R4	L	M	H
R5	VL	VH	L

The final rating is the fuzzy suitability index for each candidate route i which can be calculated by the following equation:

$$F(i) = H(i) * W_h + S(i) * W_s + RvE(i) * W_e$$

Where W_h+W_s+W_e = 1.

where W_h, W_s and W_e is the weighting factors for H(i), S(i) and RvE(i), respectively. These factors respectively depend on how the route is influenced by H(i), S(i) and RvE(i). F(i) is the approximated fuzzy number corresponding to the fuzzy index of the alternative route i. Further, we can list the final rating of the various routes to obtain the suitability index of the alternative route i.

4 PERFORMANCE METRICS

4.1 Packet delivery ratio:

It is defined as the ratio of the number of data packets received to the number of data packets sent between the source and the receivers.

4.2 Average end to end delay:

It is the average delay between the sending of the data packet by CBR and its receipt at CBR receiver.

4.3 Average route acquisition latency delay:

The average delay between the sending of a RREQ packet by a source for discovering a route to destination and the receipt of the first corresponding RREP packet.

4.4 Routing overhead:

It is defined as the ratio of the total number of control packets sent to the number of data packets delivered successfully.

rediscover the route to retransmit data packets that are lost due to the node’s mobility or unreal route paths during the communication

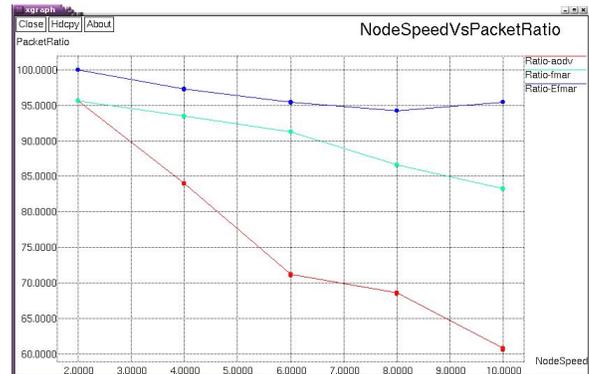


Fig 4 Node speed vs packet delivery ratio

5 SIMULATION ANALYSIS

The benefits of fuzzy modified AODV (FMAR) routing protocol can be demonstrated by the performances of the protocol with respect to packet delivery ratio, end-to-end delay, average route acquisition latency delay and routing overhead are simulated using the NS2. The FMAR functions and the lifetime of the routing agent are coded using C++ Toolkit to determine the membership functions. The mobility model used in each of the simulations is in a random direction mode. In each ad hoc network group, the nodes are initially placed randomly within a predefined 500 m * 500 m grid area. Each node then chooses a random direction between 0 and 360 degrees and may move at a speed within 0–20 m/ s. Once a node reaches the boundary of the area, it chooses a time period for remaining stationary. After the end of this pause time, the node chooses a new direction between 0 and 180. The new direction is adjusted according to the relative position toward the boundary of the area. This process is repeated throughout the simulation, and then the topology of the underlying network changes continuously. We used some applications of FTP and HTTP and a few generic CBR source with 1000 bps for generating the traffic. We simulated several rounds to change the couples of senders and receivers using topology and traffic generators. The simulation time for a round is 1500 s at the most.

5.2 Average end to end delay vs node moving speed

Fig. 5 shows the comparison of the three protocols with respect to the average end-to-end delay vs. node moving speed. It shows that the end-to-end delay increases usually with the increasing speed. It also shows that the enhancement of FMAR protocol is much better than the other two, and that the change of the delay in enhancement of FMAR is smoother than those in the other two. AODV needs more time and more control overhead than the FMAR does to recover unreal paths (broken paths) and to discover new paths.

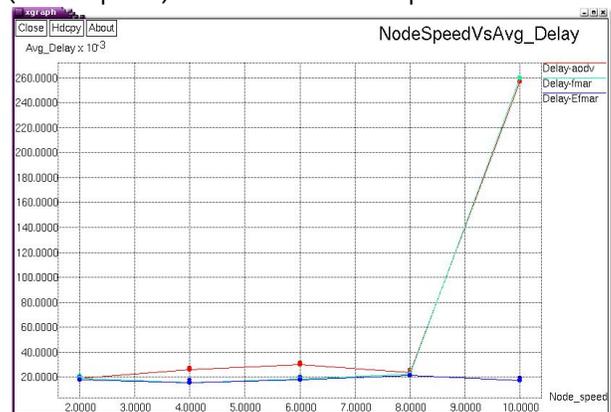


Fig 5 Average end to end delay vs node moving speed

5.1 Packet delivery ratio vs node moving speed

Fig. 4 shows the performance analysis of the achieved packet delivery ratio vs. node moving speed for the FMAR, AODV and the enhancement of FMAR protocol in an ad hoc network. packet delivery ratio is similar to throughput in that it represents the ratio of the number of data packets received to the number of data packets sent between the source and the destinations. Using AODV as a base line for comparison, the result shows that both FMAR and the enhancement of FMAR are much better than AODV protocol. This is because that AODV needs to

5.3 Average route acquisition latency delay vs. simulation time

To show that the enhancement of FMAR can find the appropriate routes in time before crashing, the average route acquisition latency delay vs. simulation time is examined. The delay is computed by noting the simulation time during the period from the time an initial RREQ or the RREQ for the re-established route are broadcast to a given destinations to the time the RREPs are received by the source. Fig. 6 shows average route acquisition latency delay vs. the simulation time. It indicates that the enhancement of FMAR protocol is also much better than

the other two with respect to the performance of the average route acquisition latency delay.

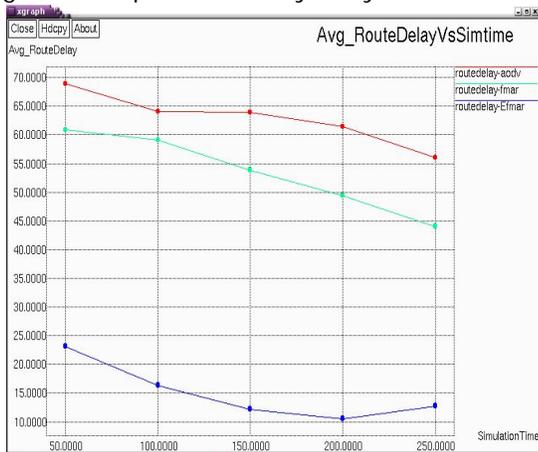


Fig 6 Average route acquisition latency delay vs simulation time

5.4. Routing overhead improvement ratio vs. simulation time

Fig. 7 shows the routing overhead improvement ratio vs. simulation time. Both the ratios are respectively obtained by dividing the overheads of FMAR and AODV based on the overhead in terms of RREQ, RREP and RERR messages sent. Since FMAR selects two comparatively stable disjoint routes for transmission instead of the quick one as AODV, therefore, the duration of routes in FMAR is better than that in AODV. When the Simulation time increase, the AODV is highly probable to use the re-established more stable routes therefore, the AODV is closer to FMAR. The enhancement of FMAR protocol is much better than the FMAR and AODV.

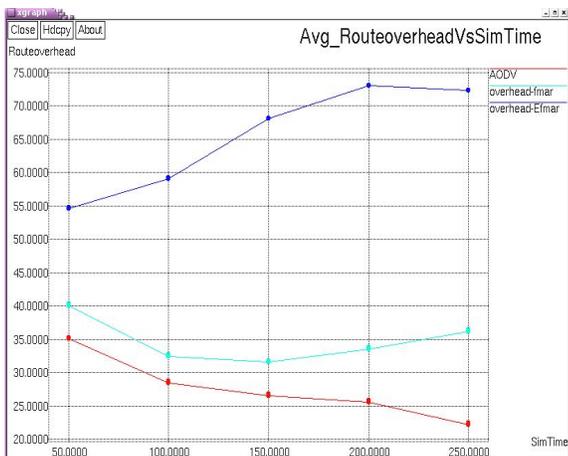


Fig 7. Routing overhead improvement ratio vs. simulation time

6 CONCLUSION

In this paper, fuzzy modified AODV (FMAR) multicast routing protocol is proposed to select two

comparably stable routes by computing dynamic route lifetime for multicast routing or layered video streaming.

The model compares and ranks different route lifetimes by the weighted Multi-criteria. The simulations of the performances of the packet delivery ratio, the average end-to-end delay, average route acquisition latency delay and routing overhead are carried out for FMAR, AODV and the enhancement of FMAR protocol. The results show that all the performances of the former two are much better than that of AODV, and the enhancement of FMAR protocol is significantly better than FMAR protocol with respect to the performances mentioned above.

REFERENCES

- [1] Bey-LingSu,Ming-ShiWang,Yueh-MingHuang. "Fuzzy logic weighted multi-criteria of dynamic route life time for reliable multicast routing in adhoc networks", *Expert Systems with Applications* pages 476- 484, 2008.
- [2] Lee, S., Su, W., etal. A performance comparison study of adhoc wireless multicast protocols. *In proceedings of IEEEINFOCOM, Tel.Aviv, Israel, 2000*
- [3] Lee, S.J., Su, W., Hsu,J., Gerla,M.,&Bagrodia,R). A performance comparison study of adhoc wireless multicast protocols. *INFOCOM*,(2000)
- [4] Agarwal,S.,Ahuja,A.,Singh,J.,&Shorey, Route lifetime assessment based routing (RABR) protocol for mobile adhoc networks. *In IEEE international conference communication* (pp.1697-1701), 2000
- [5] Ghosh,S., Razouqi,Q., Schmacher,H.J.,& Celmins,A. A survey of recent advances in fuzzy logic in telecommunications networks and new challenges *IEEE Transactions on Fuzzy Systems*, 443-447, 1998
- [6] Nasipuri, A., Castaneda, R., &Das, S.R..Performance of multipath routing for on-demand protocols in mobile adhoc networks. *Mobile Networks and Applications* 339-349, 2001
- [7] www.isi.edu/nam/ns (Ns simulator)
- [8] Perkins,C. *Adhoc networking*. Addison-Wesley,2001
- [9] Perkins,C. *Adhoc on demand distance vector (AODV) routing* Internet-Draft, draft-ietf-manet-aodv-00.txt
- [10] Wu,C.,& Tay,Y. *AMRIS: A multicast protocol for adhoc wireless networks. In military communications conference proceedings* (pp.25-29), 1999
- [11] Toh, C., Guichala, G., & Bunchua,S.ABAM: On-demand associativity-based multicast routing for adhoc mobile networks. *In proceedings of IEEE vehicular technology conference* (pp.987-993), 2000
- [12] Pasupuleti, A., Mathews, A.V., Shenoy,N., &Dianat,A. A fuzzy system for adaptive network routing. *Proceedings of SPIE* (vol.4740).Orlando, FL, USA: SPIE [1-5April], 2002
- [13] Lee,S., Gerla,M.,& Chiang,C. On-demand multicast routing protocol. *In Proceedings of IEEEWCNC1999*, New Orleans, LA.(pp.1298-1302), 1999
- [14] Paul,K.,Bandyopadhyay,S.,Mukherjee,A.,&Saha,D. Communication aware mobile hosts in ad-hoc wireless network *In Proceedings of the international conference personal wireless communication*, 1999
- [15] Chiang, T.C., Liu, C.H., &Huang, Y. M. A near- optimal multicast scheme for mobile adhoc networks using a hybrid genetic algorithm. *Expert Systems with Applications*,734-742,2007

An Investigation on The Strength and Workability of Concrete Using Palm Kernel Shell and Palm Kernel Fibre As A Coarse Aggregate

Emiero, C ,Oyedepo O.J

Abstract— The study examined the compressive strength and workability of concrete using palm kernel shell and palm kernel fibre as a partial replacement for coarse aggregate. Lightweight aggregate obtained using, Palm Kernel Shell (PKS) and Palm kernel Fibre (PKF) respectively as partial replacement for coarse aggregate. Batching was done by volume using a water cement ratio of $1.1\frac{1}{2} : 3$ and $1:2:4$. The concrete obtained of size $150 \times 150 \times 150\text{mm}^3$ was crushed using manual compression testing machine at 7, 14, 21, and 28 days respectively. The result of the specific gravity test using water displacement method and the sieve analysis for palm kernel shell are 2.59 and it has an S-shape curve; while 4.70 and 0.265 obtained for the coefficient of uniformity (C_u) and coefficient of curvature (C_c) respectively indicate that the aggregate is uniformly graded and can be used for the production of light weight concrete. Also the values of the slump obtained for mix $1:1\frac{1}{2}:3$ and Mix $1:2:4$ for concrete produced with (PKS) and (PKF) are 30mm and 38mm which indicate true slump. The concrete mix ratio PKS:PKF of 50:50 for $1:1\frac{1}{2}:3$ and $1:2:4$ has compressive strength of 12.29N/mm^2 and 10.38N/mm^2 after 28 days, which confirms light weight concrete.

Thus, from the result obtained from this research, both palm kernel shell and palm kernel fiber could be used as a replacement for coarse aggregate in light weight concrete.

Index Terms— Light Weight Concrete, Compressive Strength, Workability, Palm Kernel Shell, and Palm Kernel Fibre

◆

1.0 INTRODUCTION

The high demand for concrete in construction using normal weight aggregates such as; gravel and granite drastically reduces the natural stone deposits and this has damage the environment thereby ecological imbalance (Short and Kinniburgh, 1978), (Alengaram, Jumaat, Mahmud, 2008). Therefore, there is need to explore and to find out suitable replacement material to substitute the coarse aggregate which could be used in the production of light weight concrete.

Palm kernel shell which is found cheap in large quantities as a by-product in the production of palm oil in some parts of the country Nigeria is investigated. This is the reason why the engineering properties of cracked palm kernel shell and fibre was chosen to be analyzed so as to ascertain it suitability as a substitute for gravel/granite in production of concrete for construction.

Palm kernel shell and fibre shown in plate 1.0 are not common materials in Nigeria Construction Industry but it has found it usefulness in country such as Malaysia which is the second largest palm oil producing Country in the World and it produces more than half of World's Palm Oil (Alangaram, Jumaat, Mahmud, 2008). The requirement of vegetable oil is constantly increasing and more cultivation of palm oil is forecast in the near future (Ramli, 2003). At the same time, the production of palm oil result on by-products such as Empty Fruits Bunches (EFB), Palm Kernel Shells, (PKS) or Oil Palm Shells (OPS),

pericap, Palm Oil Mill Effluent (POME) and Palm Kernel Fibre (PKF). These are waste materials and stockpiling such wastes have caused storage problem in the vicinity of the factories as large quantities of these waste are produced everyday.

Also, these waste are stockpile in open fields, thus it has negative impact on the environment. One of the ways of disposing waste materials would be, by using them as replacement for coarse aggregate in Building and Civil Engineering Construction Works. This will help to reduce cost of natural gravel/granite and also help to prevent the depletion of natural resources and to maintain ecological balance.

Palm kernel fibres are derived from oil palm tree (*Elaeis guineensis*), (Akpe 1997, Ayanbadeto 1990) and economically valuable tree, and native to West African and widespread through out the tropics.

In Nigeria, the oil palm tree, and palm fibre are used mostly as a source of fuel for domestic cooking in most areas where they occur.

Palm kernel shell PKS is the hard endocarp of palm kernel fruit that surrounds the palm seed. It is obtained as crushed pieces after threshing or crushing to remove the seed which is used in the production of palm kernel oil (Olutoge, 1995). PKS is light and therefore ideal for substitution as aggregate in the production of light weight concrete.

Olutoge (1995) in his investigations into the physical properties of palm kernel shell found tat it

density to be 740kg/m³. He concluded that these materials has properties which resembled those of light weight concrete materials.

Olanipekun (2006) investigated the properties of palm kernel shells (PKS) as coarse aggregate to concrete. The CCS were crushed and substituted for conventional coarse aggregates in gradation of 0%, 25%, 50%, 75% and 100%. Two mix ratio (1:1:2) and (1:2:4) were used respectively. He noted that the compressive strength of concrete decreased as the percentage of the shells increased in the two proportions. His results also indicated a 30% and 42% cost reduction for concrete produced from coconut shells and palm kernel shells when used as substitute for the coconut shell were more suitable than palm kernel shells when used as substitute for conventional aggregates in concrete production. Also attempts by Abdullah (2003), Okafor (1988), and Okpala (1990) to use PKS as coarse aggregates replacing normal granite aggregate traditionally used for concrete production.

Ata et al, (2006) compared to the mechanical properties of palm kernel shell concrete with that of coconut shell concrete and reported the economy of using PKS as light weight aggregate. Generally, PKS consists of 60 – 90% particles in the range of 5 – 12.7mm (Okafor, 1988, Okala, 1990). The specific gravity of PKS between 1.17 and 1.37, while the maximum thickness of the shell was found to be about 4mm. The density of PKSC various in the range of 1,700 to 2050kg/m³ that depends on factors such as type of sand and PKS contents. Generally when the density of concrete is lower than 2000kg/dm³, it is categorized as LWC. The 28 day days cube compressive strength of about 15 – 25Mpa has been reported by them.

The present study aims to investigate the suitability of palm kernel shell PKS and palm kernel fibre PKS as replacement for coarse aggregates in the production reinforced concrete. Having carried out a brief examination of the background to this study it will be re-stated that this investigation will adopt a “waste to wealth” policy as the study materials presently have little or no economic value with disposal problem but will also ascertain their suitability as replacement for coarse aggregate in production of concrete and hence enhance their economic value.

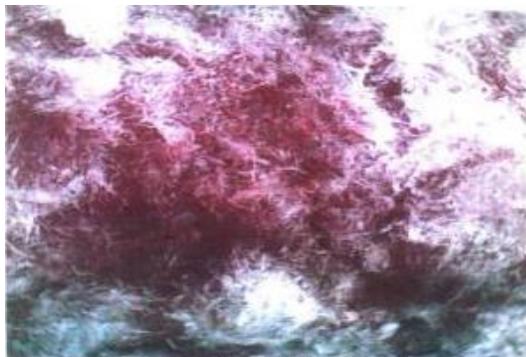


PLATE SHOWING PALM KERNEL FIBERS

Plate 1. Palm Kernel Fibers



PLATE SHOWING PALM KERNEL SHELLS

Plate 2. Palm Kernel Shells

2.0 MATERIALS AND METHODS

The materials used include the following

Sharp Sand: The sharp was source from market where they are supply in Auchi Edo State, Nigeria. The Sharp sand was sun dried to control the moisture content during usage to conform to the requirements of BS 882 (1982).

Palm Kernel Shells and Palm Kernel Fibre: These were obtained from South Ibie community in Etsako West Local Government Edo State, Nigeria. The shell and fibre were put in a basket in batches and thoroughly flushed with water to remove impurities that could be a contaminant in concrete. They were sun dried and kept in waterproof sacks.

Granite: The granite (coarse aggregate) used for the study was of range 2mm to 19mm graded, it was source from a quarry on Iyuku Auchi suburb community Nigeria, Edo State.

Cement: The cement used was ordinary Portland Cement (Dangote) it was sourced from Auchi, Edo State Nigeria and conformed to the requirements of British Standard Code (BS 12 of 1996).

2.1 DETAILS OF MIXES

The water used conformed to BS 3148 (1980), however, the water – cement ratio of 0.75 was used, while the batching was done by volume using ratio 1:1½:3 and 1:2:4.

Calculation for the volume of constituents was carried out for a target concrete strength of 15N/mm² using water cement ratio of 0.75. Volume of constituents for palm kernel shell and palm fibre replacement for batch one, as shown in table 2.0 below. Standard cone rubber bucket sizes (Depth 35cm) Diameter 30cm. (20 litres) was used as a measuring can (volumes 0.0235m²).

1.40	580	19.3	45.7
1.20	400	13.3	32.4
0.60	650	21.7	10.7
Receiver	320	10.7	0

TABLE 1
Proportion of Constituent Materials in the Concrete Mixture

PKS : PKF	CEMENT	SAND	PKS	PKF
50:50% (Mix 1:1½:3)	0.025m ³	0.0375m ³	0.0375m ³	0.0375m ³
50:50% (Mix 1:2:4)	0.025m ³	0.05	0.05	0.05m ³
GN : PKF	CEMENT	SAND	GN	PKF
50:50% (Mix 1:1½:3)	0.025m ³	0.0375m ³	0.0375m ³	0.0375m ³
50:50% (Mix 1:2:4)	0.025m ³	0.05m ³	0.05m ³	0.05m ³

2.2 CASTING OF SAMPLE

The size of from work adopted for concrete cub was 150×150×150mm. The concrete was mixed with various constituent in their respective percentage, placed and compacted in three layers after proper mixing by hand. The samples were remolded after 24 hours and kept in a curing tank for 7, 14, 21 and 28 days as required.

All tests were carried out at strength/ materials laboratory of civil engineering department Auchi Polytechnic, Auchi Edo States; table 2.1, 2.2 and 2.3 below shows the particles size distribution of the fine aggregate, coarse aggregate and palm kernel shell respectively, Figure 2.0, 2.1 and 2.2 shows their grain size distribution on the logarithmic scale. In all the mix, the crushing strength of the concrete cubes was determined using the compression testing machine. The result of workability was noted and also the crushing strength of the concrete cubes was noted 7, 14, 21 and 28 days respectively as shown in table 2.4 and 2.5 below. Each sample was weighed before putting into the crushing machine to ascertain it density. The compression strength of each sample was determined as follows

$$\text{Compressive strength} = \frac{\text{Crushing Load (N)}}{\text{Effective Area (mm}^2)} = N / \text{mm}^2$$

TABLE 2
Particles Size Distribution of the Fine Aggregate Sharp Sand.

Sieve (mm)	Mass retained (g)	%retained	% passing
6.30	0	0	100
3.35	500	16.7	83.3
2.36	300	10.0	73.3
2.00	250	8.30	65

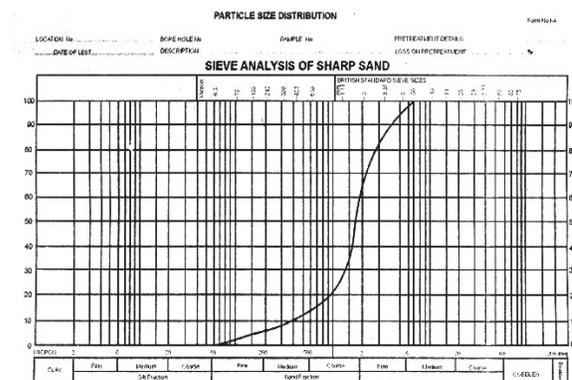


Fig. 1. The Particle Distribution Of Sharp Sand

From the graph, the uniformity coefficient *c_u* is 4.0 and coefficient of curvature or concavity *c_c* is 1.06 and the fineness modulus of sharp sand is 2.9. This means the fine aggregate is medium sand and it's good for concrete production. The specific gravity was found to 2.09. This make the fine aggregate satisfactory for production light weight concrete (M.S Shetty, 2005).

TABLE 3
Particle Size Distribution of Coarse Aggregate (Granite).

SIEVE (MM)	MASS RETAINED (G)	%RETAINED	% PASSING
24	0	0	0
19	150	6	94
13.2	350	14	80
9.53	400	16	64
6.35	600	24	40
3.35	650	26	14
2.00	300	12	2
Receiver	150	2	0

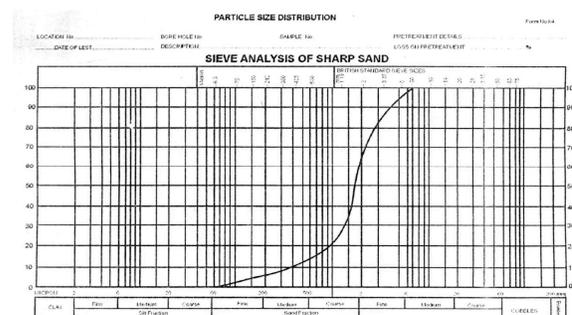


Fig. 2. Graph of particle distribution of granite

From the graph, the coefficient of curvature *c_c* is = 0.75, uniformity coefficient *c_u* is 2.4, the aggregate is

well graded. The specific gravity was found to be 2.34 which make it suitable for a production of light weight concrete and construction of road pavement (Ahmedu, A.A., 2002, Shetty, M.S., 2005)

TABLE 4
Particle Size Distribution of Palm Kernel Shell

SIEVE (MM)	MASS RETAINED (G)	%RETAINED	% PASSING
19	0	0	100
13.2	300	10	90
9.63	750	24	65
9.50	400	16.4	52.6
6.35	900	30.0	22.6
3.35	430	14.3	8.3
Receiver	250	8.30	0

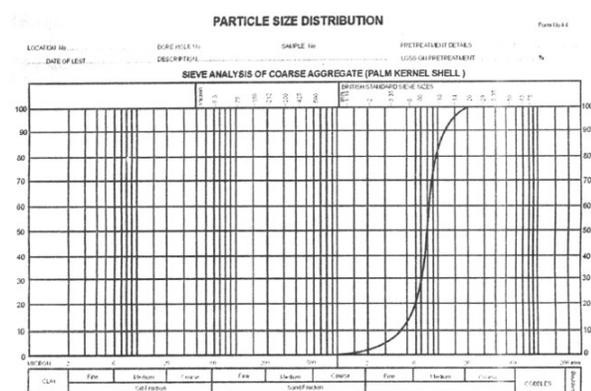


Fig. 3. Particle Size Distribution of Palm Kernel Shell

From the graph, the uniformity coefficient is 4.70 and coefficient of curvature $cc = 0.265$. It means the aggregate is well graded, can be used for production of light weight concrete. Also the specific gravity was found to be 2.29 which make it suitable for light weight concrete, (Ahmed, A. A. 2002, Shetty, M. S. 2005). The palm kernel shell impact value was found to be 50% which confirmed its suitability for light weight concrete production of grade C10 to C15 but weak for pavement construction (Ahmed, A. S. 2002).

2.3 TEST ON CONCRETE

TABLE 5
Summary of Workability Test on the Various Mix Ratios

Mix Ratio	Slump (mm)	Compaction Factor Value	Remark
PKS : PKF 1:1½ : 3	30	0.90	True slump
PKS : PKF 1 : 2 : 4	38	0.70	True slump

GN : PKF 1 : 1½ : 3	28	0.94	True slump
GN : PKF 1 : 2 : 4	34	0.90	True slump

From the slump and compaction result of the various mixes the concrete made is of high degree of workability. (Shetty, M. S. 2005).

TABLE 6
Crushing Strength Value of Concrete Work

Cube Ref	Mix Ratio	Weight	Days /Age	Fracture Load (KN)	Compressive Strength (N/mm ²)	Density (Kg/m ³)
PKS:PKF 50:50%	1:1½ :3	5.70	7	109.12	4.85	1688.90
	1:2:4			109.20	4.85	174.10
PKS:PKF 50:50%	1:1½ :3	6.10	14	160.00	7.04	1807.40
	1:2:4			151.12	6.27	1866.70
PKS:PKF 50:50%	1:1½ :3	6.30	21	181.83	8.08	186.70
	1:2:4			161.28	7.44	1926.0
PKS:PKF 50:50%	1:1½ :3	6.60	28	276.37	12.29	1958.6
	1:2:4			243.43	10.83	2014.60
GN:PKF 50:50%	1:1½ :3	6.70	7	72.72	3.23	1985.2
	1:2:4			87.12	3.87	2044.40
GN:PKF 50:50%	1:1½ :3	6.90	14	145.44	6.47	2044.40
	1:2:4			159.84	7.10	213.30
GN:PKF 50:50%	1:1½ :3	6.80	21	196.84	8.71	2014.80
	1:2:4			218.16	9.70	2044.40
GN:PKF 50:50%	1:1½ :3	7.0	28	232.93	10.35	2074.0
	1:2:4			261.73	11.63	2192.60

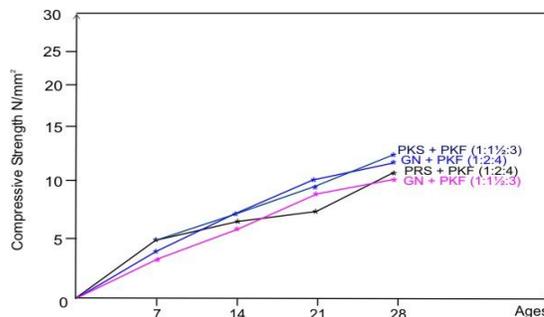


Fig. 4. Compressive Strength of Concrete Mixes

3.0 EFFECT OF REPLACEMENT OF COARSE AGGREGATE WITH PKS PLUS PKF IN CONCRETE

From the table, the maximum compressive strength of palm kernel shell plus palm kernel fibre is 12.29N/mm² and 10.83N/mm² which confirm light weight concrete strength of 10 to 15N/mm² and 10.35N/mm² was achieved when granite plus palm fibre was used for mix 1:2:4 having the highest strength against 1:1½:3. This also confirm light weight concrete strength although didn't follow the conventional

strength attainment of the highest mix ratio having the highest strength. This could be as a result of reaction of fibre with concrete cementations materials.

It was observed that rate of absorption for water increase from 7 days to 28 days about 9.2% for PKS plus PKF for mix 1:1½:3 while mix 1:2:4 is 13.0%. It can be seen also that the percentage absorption of water for GN plus PKF of mix 1:1½:3 is 4.3% while mix 1:2:4 is 6.8%. This shows that PKF plus PKS concrete absorbed more water compared to concrete made with replacement of granite with PKF.

4.0 CONCLUSION/REMARK

From the result of the test carried out:

- i. The possibility exists for the partial replacement of coarse aggregate with palm kernel shell plus palm kernel fibre to produce light weight concrete as recommended by BS: 8110, (1997).
- ii. With 50:50 replacements of coarse aggregate with GN plus P.K.F. The strength attainments vary from mix 1:2:4 to 1:1½:3 with the later having the highest strength of 12.29N/mm².
- iii. With 50:50 replacement of coarse aggregate with GN plus P.K.F. The strength attainment varies from mix 1:1½:3 to mix 1:2:4 with the later having the highest strength of 11.63N/mm². It indicates that granite is better compared to all other natural available aggregate.
- iv. Organic materials are subject to deterioration over time hence concrete made with PKS + PKF and GN + PKF should be regularly maintained and replaced when necessary.
- v. From its density which ranges between 1688kg/m³ to 2014kg/m³ for concrete made with PKS plus PKF; it can be used only as light weight concrete, while partial replacement of granite with palm kernel fibre with a density running between 1985kg/m³ to 2192kg/m³ can be used as light weight concrete as well as dense concrete if the proportion of palm kernel fibre (PKF) is reduced to 10 to 25%.

5.0 REFERENCES

- [1] Akpe E. (1997) : Characteristics Strength of Concrete Using Palm Kernel Shell As Light Weight Concrete Aggregate, Unpublished Technical Report, Department of Civil Engineering EDUSUL Ekpoma, Edo State.
- [2] Ata.O, Olanipekun E.A. and Olusola k. O. (2006): A Comparative Study of Concrete Properties Using Coconut Shell And Palm Kernels Shell As

Coarse Aggregate Journal of Building And Environment AL (3):297-301)

- [3] Ayanbadefo, S. O. (1990): Investigation Into The Use Of Palm Kernel Shell As Light Weight Aggregate For Concrete. Unpublished Thesis, Department of Civil Engineering, University of Benin.
- [4] Jumaat. A. Alengaram, U. J. (2008): Comparison of Mechanical and Bond Properties of Palm Kernel Shell: Journals: www.academicjournals.org/ijps/pdf
- [5] Okpala, D.C.I (1990): Comparative Study of Concrete Properties using Coconut Shell and Palm Kernel Shell as Coarse Aggregate , Journal of Building and Environment 41(3): 297-301;
- [6] Olanipekun, E. A. (2006) A Comparative Study of Concrete Properties Using Coconut Shell And Palm Kernel Shell As Coarse Aggregates, Journal of Building And Environment 41(3): 297-301.
- [7] Olutoge F.A (2010): Investigations on Sawdust And Palm Kernel Shells As Aggregate Replacement, ARPN Journal Of Engineering And Applied Sciences VOL.5. NO4, April 2010.
- [8] Olutoge, F.A (1995): A Study of Saw Dust, Palm Kernel Shell And Rice Husk Ash As Full/Partial Replacement For Sand, Granite, and Cement In Concrete: Unpublished M.Sc Thesis, Department of Civil Engineering, University Of Lagos Nigeria
- [9] Ramli R. B. (2003): Journal on International Investment and Free Economic and Social Research, Faculty of Economics, University of Indonesia;
- [10] Short .A. and Kinniburgh (1978): Ductility behavior of Reinforce Palm Kernel Shell Concrete Beams:www.eurojournals.com/ejsr
- [11] Short A. and Kinniburgh, W. (1978): Light Weight Concrete Materials properties, specification aid testing, Noyes publications: www.academicjournals.org/ijps/pdf Applied science London.

Personalizing M-learning Systems Using Recommender Systems

Mahsa Baniardalan and Mehregan Mahdavi

Abstract—In mobile learning, there is no limitation for the number of learners; it is more attractive than traditional learning, more convenient and cheaper for updating, quicker in terms of broadcasting, less bulky and it is easier to transfer information. In addition, the studies show that today, personalization is an important prerequisite for attracting learners in order to consider their requirements and interests and to provide proper context for the learners. This can result in improving the quality of learning. In this paper, a comprehensive model is presented which includes personal, contextual, and environmental properties. Based on these parameters, the recommender system suggests educational materials in a proper manner such that the learner is attracted to the learning system. The system's performance could be evaluated using questionnaires.

Index Terms— Content, Context Aware, Mobile Learning, Personalization, Recommender Systems.

1 INTRODUCTION

At the present era of information technology, it is possible to have a mobile form for every application, such as is learning. The category of mobile learning is a new concept which has been introduced after electronic education has been developed in order that learning and training can be available in everywhere and every time and it can be evaluated as well.

Mobile learning has been given various definitions including: Mobile learning can be considered as a combination of "distance education" and "electronic" because the trainer and learner are separated by a distance and the learning is facilitated by advanced technologies using electronic tools [1]. According to another definition, mobile learning is an introduction to ubiquitous learning in which the learner can access to his or her required information everywhere and every time [2].

As the whole, this method of learning can be considered as everywhere and every time learning through mobile tools such as portable computers and mobile phones, with no need for physical connection to the network [1, 4].

Mobile learning can provide effective methods of learning, specifically educational courses for those who are not able, for any reason, to attend in full time classes. Mobile learning has been successful in Europe and Africa [3].

In another definition, mobile learning is defined as any learning in which learner is not in a preplanned and stationary place and the learning is processed through mobile technology [5]. In other words, mobile learning does not depend on a specific place and it is flexible in terms of place of learning and mobile technology. In fact, the mobility of learning is on focus [17]. Therefore, the importance of mobile learning is that it makes possible to access learning everywhere [6]. In addition, it is interactive and enjoyable and it is easily used to learn efficiently and more amusingly.

The previous research has shown that educational outcome is not the same for all people and it often fails in spite of spending much money and consuming a lot of time. Users and learners in these systems differ in terms of their previous knowledge, learning achievement rate, age group, skill and profession and personal intention to learn [7]. If mobile educational system is designed based on the user properties, the quality of electronic education will be improved.

This research presents a simple but efficient model for personalizing mobile learning in order to meet all needs of learners.

This paper is organized as follows: Section 2 presents content analysis and the way the required content is developed. It also presents existing models. An efficient model is introduced in Section 3, which considers different aspects and priorities of learners in order to provide learners with suitable educational materials according to their needs. Finally, Section 4 concludes the paper.

2 LITERATURE REVIEW

2.1 Recommender Systems

A recommender system in electronic education recommends

• Mahsa Baniardalan, Department of Information Technology, International Campus, University of Guilan, Iran, E-mail: mahsa_ardalan@yahoo.com

• Mehregan Mahdavi, Department of Computer Science and Engineering, University of Guilan, Iran, PH-01316690274. E-mail: mahdavi@guilan.ac.ir

the optimal content to learners. In recent years, a lot of studies have been conducted on the process of guiding users in educational systems [11, 13]. Generating useful recommendations to a group of users for items or products that might interest them is the main purpose of recommender systems. Some examples of operation of industry-strength recommender systems are suggesting books on Amazon, or movies on Netflix.

Most of the time, the so-called information overload problem makes a lot of difficulties for users to locate the right information at the right time. Recommender systems have come out as an important answer to it. It has been completely spread out in variety of applications, mostly in the form of intelligent virtual assistants in a variety of e-commerce domains. Recommender systems have demonstrated to be effective at delivering the user a more intelligent and proactive information service. They use techniques borrowed from user profiling, information filtering and machine learning, in order to make proper recommendations for products or services based on users' preferences and needs [14].

Generally two recommendation strategies are more important. One of them is Content-based recommenders which rely on rich content descriptions of items (products or services) that are being recommended [15]. A content-based movie recommender which typically rely on information such genre, actors, director, producer is an example of and matches these criteria against the learned preferences of users in order to select a set of promising movie recommendations. Clearly, the design of such a content-based system should perform considerable knowledge-engineering activities since the required domain knowledge may not be readily available or easily extracted. The collaborative filtering (CF) recommendation strategy takes advantage of the fact that in real world we often look at our friends for recommendations. It does not need item knowledge and instead uses user profiles which also include the past rating history or behaviour of the users. It is also called social filtering. By using ratings history of a set of suitable users, it can generate useful recommendations for a target user. The set of users that are chosen by the system are chosen based on the similarity of the ratings history with the target user [14].

2.2 Context Definition

Context is defined as a collection of all information about a certain entity [10]. In mobile learning, the collection of information is related to properties and situation of learners which can be collected explicitly (asking directly from the user) or implicitly (the system draws out what it needs secretly from the user) based on the behavior of users and suggests the best alternative to the learners [18].

In fact, any information related to improved interaction among learners, software and environment is called context [8]. The learners' needs are often latent and mostly they can be identified from habit, morals and even from learners' motivation. To decrease the interaction challenges, the system designer must identify learner properties to know how the learner can be attracted to the system and how he or she can constantly attempt learning [9].

2.3 Personalization Process

Personalization is based on the modeling of learning context, in a collection of information that characterizes the situation and environment in which the learner is. It comprises the information and assumption about the learner (such as personal profile, goal, knowledge, interest preferences, interaction, and presentation history) and the information about the environment (such as location, devices, time, date and weather) [11].

It is clearly recognized that the personalization process is essential. Each person's properties differ so his or her capabilities vary and we have to consider these properties. Personalization process is implemented using the identified context related to the learner. The purpose is to suggest a packet of activities to the learner based on his or her priorities and interests such as links, articles, materials, tests, homework and so on [10].

By personalization based on learning context, we mean that certain learner-related information such as goal, situation, environment and assumptions are gathered and categorized as described in the followings.

Various definitions are previously provided for context. For instance, primary types of context can be considered as location, identity, time and activity [18]. Context can be considered in different categories. Computing context include information such as: network connectivity, communication costs, communication bandwidth. Information such as users' profile and location create user context. There are also information on the environment that create environment context, such as noise levels, lighting situation, and temperature [19]. Moreover, attributes of people and objects around a user and the changes in such information can be considered as context [20, 21]. Context-aware systems are defined as systems that are able to sense the context and act upon available context information [22]. In a different categorization of context it can be considered as i) environment (physical and social), ii) self (device state, physiological and cognitive), and iii) Activity (behavior and task) [23]. In other words, context can be defined as any information that can characterize the situation of an entity (e.g. person, place, or object) [24, 25]. Context can also be categorized as externally-imposed, externally-induced or internally-induced [26]. An ontology-based context model may consider time, place, user knowledge, user activity, user environment and device capacity [27]. In some previous research, a situation model was defined to give a view on the context model describing temporal properties. It used the following context parameters: variety, priority, granularity, implementation, and cost-effectiveness [28]. Data gathering techniques about the context is an essential issue in pervasive computing [29]. Moreover, a previous research in mobile learning has argued that a context aware mobile learning system should also include the learner's willingness to participate in the proposed learning activity [30].

3 THE PROPOSED MODEL

In this model, as it is illustrated in Figure 1, assumptions and information about learners are categorized into many parts.

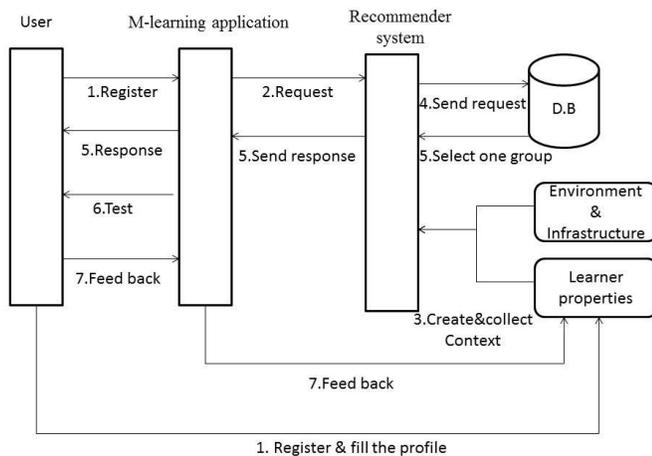


Figure 1. The proposed model for mobile learning recommender system

The recommender systems development process includes the followings:

1. Register: There are two modes. In the first mode, an intended user for learning, registers in his or her application and fills the questionnaire in the system and gives the primary information. In the second mode, the content is sent according to the learner's needs. Learner's profile is empty and it is gradually filled implicitly in the system or partly filled by the learner in order for the system to identify the learner's interests. The studies show that learner's priorities are static. However, it is possible that the priorities change. We can permit the system to select automatically for the learner and overcome this challenge.
2. The learner's application is sent to recommender system (of course, it can be spontaneous in certain cases and the information is sent to learner without primary application).
3. Content creation and collection: The information is collected in various ways such as the followings:
 - The learner defines certain variants such as demographic data and level of education, favorite orders such as using blue color for the texts or using audio or video. His or her field of interest (math, science and so on), his or her goal of learning (practical use or professional use), physical disabilities in terms of hearing, weight, social behavior such as if the learner is introvert or extrovert, whether the learner thinks positively or negatively, what is the learner's motivation to learn, what about his or her feelings and intelligence index, and so on.
 - Certain variants can be identified by the trainer (learning theory).

- Some variants can be identified by the parents (expenditure limit and security measures).
- Variants which are identified by partnership companies (evaluations).
- Variants which others such as physicians, psychologists, scholars, friends, producers and suppliers in communication networks identify.
- Pretests which are given to learners to evaluate and achieve certain variants (such as learning style and learner's level).
- Some variants are achieved through software and hardware systems (such as health, weather, location of learner and maps).
- Data mining techniques can develop learner related knowledge (such as interests, media exchange, equipment feature) or using online or offline database such as social networks, state database.
- Computer producers can monitor certain variants to build up data (such as current situation, feelings and emotions, results and achievements).
- Tracking using sensor, GPS and wireless sensor network (WSN) or radio frequency identification (RFID).

User personal properties and environment data can be obtained by these variants.

4. To request user's properties from database with various formats such as text, audio, video, animation and so on which are collected and categorized by the experts for different groups.
5. To send the materials from database to recommender system and through mediatory to the intended user.
- 6,7. After the user finishes the learning, the system examines the learner spontaneously by self test approach and the result is stored in user personal profile in order for the user know what should be sent during the next session (if learner has learned well, more advanced materials will be received and if it should be repeated, the previous material would be explained for the learner).

The compoments of the proposed recommender system include the followings:

1. User (learner): The one who uses system and helps to improve it.
2. Mediator: The operational system which performs software and plays as an interface between the user and recommender system to facilitate their relations.
3. Recommender system: It contains three inputs and a collection of best recommendations for the user as the output. The total goal and system assessment and whether it is suitable or not will depend on the output.
4. Database: It contains related educational information for user personal profile and priorities which is designed and categorized by the experts and is presented to the user based on the user needs. Of course, for a real time and secure system, a high memory and high secure database is required. In addition, learning contents must be also modified before being sent to the receiver's

device by agents or filters that made the format transformation and serve contents adapted to them. This filter should act depending on the type of devices [16]. The contents can be classified as: Text, Image, Audio and video

- Text: Text can be easily converted to another mobile device's adapted format, but, due to the increase on e-book readers, there are a lot of new standards that modify the pagination and font size automatically. These make text easily readable on those devices. Among these new formats, the most used are EPUB and PRC and LIT.
- Image: Although most of the image formats can be watched properly on mobile devices, it is useful but not essential, to scale them or to modify them before being sent.
- Audio: Mobile devices, ipod/iphone and MP3 players support most of the audio formats, so a recording is not necessary. However, it will be necessary in case of video stream.
- Video: For a good visualization, the video format must be transcoded [17] to be adapted to the basic mobile devices format as MP4 for iphones or ipods, and 3GP for rest [18].

5. User environment which includes:

- Location: It is very important for applications, especially in location based learning, to use WSN, RFID and GPS technologies.
- Tools: Mobile tools play the main role for presentation and projection such as screen size, power of processing, data store, connection method, battery consumption.
- Time: The type of content depends on the intended time
- Weather: The type and framework of content depends on the environment condition, crowded or not crowded, rainy or not rainy, cold or warm.

6. Learner's properties: These properties may include:

- Personal profile: Personal information of learner is used to classify the learner in a proper category and to assign him or her suitable learning content.
- Goal: Learners have a specific goal when they use a learning system. By identifying their goals, the system can supply the proper content for the learner. So, it can be regarded as a filter which is able to classify the learners with the same goal.
- A history of interactions and presentations: It is useful in terms of: (i) The learner is aware of what he or she has learned (for example, the learned books can be colored blue), and (ii) the personalization process can be improved by the system.
- Knowledge level: It seems that it is the most important role of learning system and can be measured by scalable tests in terms of quality (in three category: beginner, middle and advanced) and quantity (0-10). The system does not have any information about the learner so it updates know-

ledge model constantly. For example, when a learner learned the material, the system examines and according to the results, it raises or decreases the knowledge level (knowledge propagation process).

- Interests: They are categorized as well and if the learner shows interest towards the content, it will be explained in detail and if not, the content is overviewed. It is noted that at first, the interest rate is null and it is gradually improved and updated.
- Priorities: The learner can set up what is presented and how it is presented (page by page, as slides or all-in-one). The alternatives can be various media such as audio, video, animation, text and so on which can be presented by spontaneous selection in the system or by directly selection of the learner.

Figures should be numbered consecutively as they appear in the text.

4 CONCLUSION

In general, learners can learn using various tools, in different environments and through many networks with diverse levels of learning and interests. Learner must be supported by adjustable recommender systems. The suggested model considers the entire user's needs and tailors the content properly for the learner using the database of the system. But the proper feedback is implemented by post-tests. Although the previous models considered certain useful details, in the present paper, we attempt to design a model which incorporates entire needs of the user and suggests the proper content in order to meet the needs of user and raise the satisfaction of learner.

REFERENCES

- [1] Safarzade, M. and Manuchehri, K., "Learning Via Mobile Electronic Devices", 2nd International Conference Electronic Municipality, Tehran, Irans Interior ministry, 2009.
- [2] Yazdi, J. and Kahani, M., "Resource Management in a Semantic Mobile e-Learning Grid", 12nd International Conference on Computer Society of Iran, Tehran, Iran, 2006.
- [3] Abdolahi, M., "Mobile learning and role of Future of E-learning", 2nd Conference on E-learning, Zahedan, Iran, 2007.
- [4] Ataran, M., "Curriculum prospective." Monthly School Tomorrow. Five years.
- [5] O'Malley, C., vavoula, G., Taylor, J., sharples, M., Lefrere, P., Lonsdale, P., Nismith, L. and Waycott, J., "D4.1 Guidelines for learning in a mobile environment", MOBlearn, UoB. OU, wp, d4.1/1.2. 2005.
- [6] McFarlane, A., Roche, N. and Triggs, p., "Mobile learning: research finding. Coventry", Becta. [online] available: http://partners.becta.org.uk/uploaddir/downloads/page_documents/research/mobile_learning_july07.pdf, 2007
- [7] Jane, Y-K and Joy, M., "A context-aware personalized m-learning application based on m-learning preferences", International Journal of Mobile Learning and Organisation, Vol. 5, No. 1, pp. 1-14, 2010.
- [8] Jappinen, A., Nummela, J., Vainio, T. and Ahonen, M., "Adaptive Mobile Learning Systems-The Essential Questions from the Design Perspective", In the Proceedings of MLearn2004, Roma, Italy, 2004.

- [9] Yudelso, M., Gavrilova, T., and Brusilovsky, P., "Towards user modeling meta-ontology", Proc. of 10th International User Modeling Conference, Springer Verlag, pp. 448-452, 2005.
- [10] Economides, A., "Adaptive context-aware pervasive and ubiquitous learning", International Journal of Technology Enhanced Learning, Vol. 1, No 3, pp. 169-192, 2009.
- [11] Zhou, R., and Rechert, K., "Personalization for location-based E-learning", The Second International Conference on Next Generation Mobile Applications, Services, and Technologies, IEEE Computer society Washington. DC, USA, pp. 247-253, 2008.
- [12] Dunn, R., & Dunn, k., "Teaching students through their individual learning style: A practical approach", Reston.VA: prentice-Hall, 1978.
- [13] Uday Bhaskar, N., and Govindarajulu, P., A design methodology for acceptability analyzer in context aware adaptive mobile learning systems development. IJCSNS International Journal of Computer Science and Network Security, Vol. 8, No. 3, pp. 130-138, 2008.
- [14] O'Donovan, J., Smyth, B., "Trust in recommender system", in proceeding of 10th international conference on intelligent user interfaces, Acm press, California, USA, pp. 167-174, 2005.
- [15] Melville, P., Mooney, R., and Nagarajan, R., "Content boosted collaborative filtering in improved recommendation", Proceedings of the Eighteenth National Conference on Artificial Intelligence(AAAI-2002), Edmonton, Canada, July 2002.
- [16] Houser, C., and Thornton, P., "Poodle: a course-management system for mobile phones". Wireless and mobile technologies in education, Proceedings of the third IEEE International Workshop on Volume, Los Alamitos, California, IEEE Computer Society, pp. 215-215, November 28-30, 2005.
- [17] Reveiu, A., Dardala, M., Furtuna, F., "Using content multimedia data retrieval for multimedia content", Proceedings of the 12th Int. Conf on Human-Computer Interaction HCII, Vol.3 LNCS 4552: Lecture Note in Computer Science, pp. 486-492, 2007.
- [18] Schilit, B.N., Theimer, M.M. and Welch, B.B. "Customizing mobile applications", Proceedings of the USENIX Symposium on Mobile and Location-Independent Computing, (USENIX Association), pp. 129-138, 1993.
- [19] Schilit, B.N., Adams, N. and Want, R., "Context-aware computing applications", Proceedings of IEEE Workshop on Mobile Computing Systems and Applications, Santa Cruz, California, 1994.
- [20] Becker, C. and Durr, F., "On location models for ubiquitous computing", Personal Ubiquitous Computing, Vol. 9, Issue 1, pp. 20-31, 2005.
- [21] Schilit, B.N. and Theimer, M.M., "Disseminating active map information to mobile hosts", IEEE Networks, Vol. 8, No. 5, pp. 22-32, 1994.
- [22] Ryan, N., Pascoe, J. and Morse, D. "Enhanced reality fieldwork: The context-aware archaeological assistant", Computer Applications in Archaeology, British Archaeological Reports, Oxford, 1998.
- [23] Schmidt, A., Beigl, M. and Gellersen, H.W., "There is more to context than location", Computers and Graphics Journal, Vol. 23, No. 6, pp. 893-902, 1999.
- [24] Dey, A.K. and Abowd, G., "Towards a better understanding of context and context-awareness", Proceedings 2000 Conference on Human Factors in Computing Systems (CHI), The Hague, The Netherlands, 2000.
- [25] Tamminen, S., Oulasvirta, Toiskallio, K. and Kankaninen, A., "Understanding mobile contexts", Personal and Ubiquitous Computing, Vol. 8, Issue 2, pp. 135-143, 2004.
- [26] Hill, J.R. "Reflection on resource-based learning environments: Continuing the exploration of opportunities and obstacles", International Journal of Knowledge and Learning, Vol. 3, No. 1, pp. 12-29, 2007.
- [27] Bouzeghoub, A., Do, K.N. and Lecocq, C., "A situation-based delivery of learning resources in pervasive learning", E.Duval, R. Klamma, and M. Wolpers (Eds.): EC-TEL 2007, LNCS 4573, pp. 450-456. Springer-Verlag Berlin Heidelberg, 2007.
- [28] Bayoumi, F., "Context aware systems: Present and future", Proceedings of the IASTED European Conference, Internet and Multimedia Systems and Applications, Chamonix, France, pp. 208-213, 2007.
- [29] Roibas, A.C. "Teaching appropriate ethnographic methodologies for pervasive computing", International Journal of Teaching and Case Studies, Vol. 1, No 1/2, pp. 135-145, 2007
- [30] Bhaskar, N.U. and Govindarajulu, P., "A design methodology for acceptability analyzer in context aware adaptive mobile learning systems development", International Journal of Computer Science and Network Security, Vol. 8, No. 3, pp. 130-138, 2008.

Study of Ion Sound Propagation Waves in Thermal Plasma

Ghoutia Naima Sabri^{1*}

Abstract— Plasma can support a great variety of wave motion. Both high frequency and low frequency, electromagnetic and electrostatic waves may propagate in plasma. The primary emphasis has been placed on the study of electrostatic waves because the ease with which such waves may be excited and detected and because the collisionless damping of waves predicted by Landau can be conveniently studied. In this paper we will discuss the propagation of ion sound waves in the collisionless thermal plasma by calculating the dispersion relation using fluid theory. The kinetic treatment shows clear that these waves are subject of strong Landau damping for weak temperature of ions compared to the one of electrons. The interest of this study can be applied to some astrophysical phenomena more precisely in the study of the generation of these waves in the topside ionosphere at low latitude sunrise and sunset.

Index Terms— Ion sound waves, thermal plasma, dispersion, fluid, kinetic, Landau, damping, temperature.

1 INTRODUCTION

IN Plasmas, there are a variety of waves which propagate with a variety of nonlinear dispersion relations. We will examine one type of these waves and the approximations needed to find its dispersion relation. Ion sound wave is a longitudinal electrostatic wave in unmagnetized plasma arises from the motion of the ions by assuming that the frequency is low enough that ion can participate in motion. For small wave number k , it has the linear form of a normal sound wave. In this study we will discuss the propagation of these waves in non-collisional plasma, taking into account thermal effects. Thermal changes in wave propagation are not well described by fluid equations. To do this we will use the kinetic description of plasma and the appropriate equation is the *Vlasov equation*.

In hot plasma, the dispersion function and its derivatives have a wide range of applications in the descriptions of waves of small amplitude. It is also widely used in the description of polarization strongly inhomogeneous media. Accurate assessment of this function is important in various fields of science. *Fried and Conte* (1961) have presented interesting work on the main properties of the dispersion function of hot plasma.

2 BASIC PROPERTIES OF HOT PLASMAS (THERMAL), IONISATION, SAHA LAW

In thermal plasmas, collisions between particles can cause ionization if the energy difference between the particles is enough large (of the order of a few eV), or the recombination, if the energy difference is quite low.

As in the same ionized gas, the two forms of collisions can occur, a balance can be established. Just to maintain this balance that the plasma is hot enough. It must even have a temperature of several tens of thousands of degrees [1], [2] eg stars and nuclear explosions.

The state of ionization of plasma is related to its temperature T and density n and the degree of ionization which is defined by

$$\alpha = \frac{n_e}{n_0 + n_e} = \frac{Z n_i}{n_0 + Z n_i} \quad (1)$$

Where n_e is the electron density, n_i the ion density and n_0 neutral density. Because of collisions, atoms, molecules, or ions in the plasma can be ionized if the temperature is such that

$$k_B T > U_i / 10 \quad (2)$$

U_i is the ionization potential. If plasma is in thermodynamic equilibrium, the ionization is balanced by recombination. This balance is described by the *Saha equation* [3].

$$\frac{n_i}{n_0} = \frac{(2\pi m_e k_B T)^{3/2} T^{3/2}}{h^3} \frac{1}{n_i} \exp[-U_i / k_B T] \quad (3)$$

Where n_0 is the neutral density, n_i the density of ions et $n_i = n_0$ is the balance between the ionization rate (depending on T) and the rate of recombination (depending on density) h is Planck's constant ($h = 6.62 \cdot 10^{-34} \text{ J.s}$) and $(2\pi m_e k_B T)^{3/2} / h^3$ correspond to the thermal wavelength of an electron ($\lambda = 2.4 \cdot 10^{11}$).

The term that contributes the most is $\exp[-U_i / k_B T]$

- If $U_i \gg k_B T$ low ionisation, $\alpha \rightarrow 0$ (industrial plasmas and ionosphere).

- If $U_i \ll k_B T$ high ionisation, $\alpha \rightarrow 1$ (thermonuclear plasmas and stellar).

Typically, α begins to be meaningful when $k_B T > U_i = 10$ and allows distinguishing between weakly and strongly ionized plasmas.

*Ghoutia Naima Sabri is doctor in science physic, lecturer in University of Bechar, Algeria, PH-0213773146965. E-mail: sabri_nm@yahoo.fr

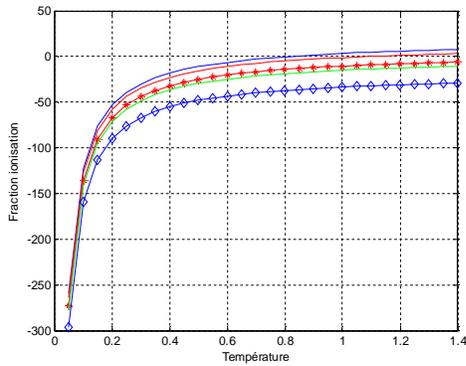


Fig.1. (a) The ionization of a fraction of hydrogen ($U_i = 13.6 \text{ eV}$) in function of temperature and density (log scale).

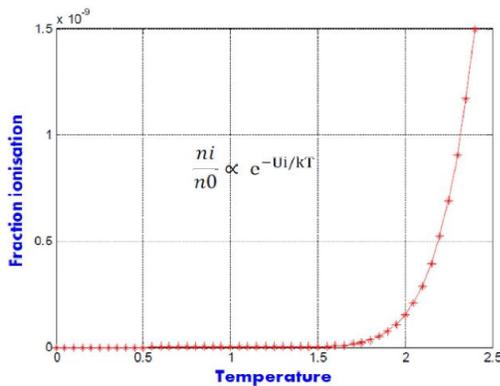


Fig. 1. (b) The ionization of a fraction of hydrogen ($U_i = 13.6 \text{ eV}$) for a constant density (Saha law).

3 VLASOV MAXWELL SYSTEM

The analysis of the behaviour of particles moving in a hot plasma is based on the Boltzmann equation of the function of distribution $f_s(\vec{r}, \vec{v}, t)$, also known as the Vlasov equation [4], [5]. This equation characterizes the evolution in time and space distribution of particles of non-collisional plasma in kinetic description. For the species s , we can write a kinetic equation of Vlasov in the form

$$\frac{\partial f_s}{\partial t}(\vec{r}, \vec{v}, t) + v \cdot \nabla_r f_s(\vec{r}, \vec{v}, t) + \frac{q_s}{m_s} (\vec{E} + \vec{v} \wedge \vec{B}) \cdot \nabla_v f_s = 0 \quad (4)$$

This latter which is coupled to the Maxwell equations allows to describe the evolution of electric and magnetic fields. In the presence of electromagnetic field (\vec{E}, \vec{B}) , Maxwell's equations are

$$\nabla \cdot \vec{E} = \frac{1}{\epsilon_0} \sum_s q_s \int f_s d^3 \vec{v} \quad (5)$$

$$\nabla \cdot \vec{B} = 0 \quad (6)$$

$$\nabla \wedge \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (7)$$

$$\frac{1}{\mu_0} \nabla \wedge \vec{B} = \epsilon_0 \frac{\partial \vec{E}}{\partial t} + \sum_s q_s \int \vec{v} f_s d^3 \vec{v} \quad (8)$$

The system consists of the Vlasov equations (4) and Maxwell (5) - (8) is closed and called Vlasov-Maxwell system. In the absence of magnetic field applied from outside, the field \vec{E} will be zero (non relativistic case), the isotropic medium is called electrostatic and Lorentz force is reduced to an electrical force $q_s \vec{E}(\vec{r}, t)$ and the system of Vlasov-Poisson [6] and the equation is written

$$\frac{\partial f_s}{\partial t} + v \cdot \nabla_r f_s + \frac{q_s \vec{E}}{m_s} \cdot \nabla_v f_s = 0 \quad (9)$$

4 DISPERSION FUNCTION IN A HOT PLASMA

It is easy to see that, according to an appropriate scale variable as the dispersion relation for electrostatic waves is expressed by the dispersion function Z :

$$Z(\eta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{e^{-\xi}}{\xi - \eta} d\xi, \text{Im}(\eta) > 0 \quad (10)$$

Moreover, $Z(\eta)$ is the Hilbert transform of a Gaussian [7],[8]. With $\eta = \frac{\omega}{v_{Th}}$, $\xi = u/v_s$ and $v_{Th} = \sqrt{2k_B T_s/m_s}$ and $v_{Th} = \sqrt{2k_B T_s/m_s}$

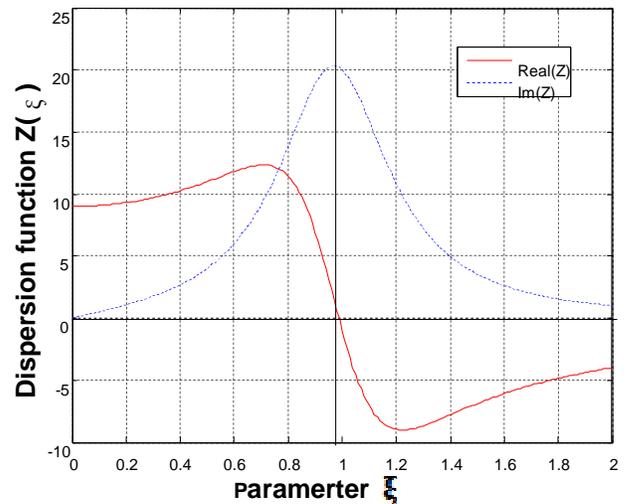


Fig. 2. Real (full curve) and imaginary (dashed curve) parts of plasma dispersion function $Z(\eta)$

As is shown on the figure 2, for $\text{Im}(\eta) > 0$, this function is defined in the upper half complex plane, the analytic continuation in the lower half-plane is obtained by writing

$$Z(\eta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{e^{-\xi}}{\xi - \eta} d\xi + i\sqrt{\pi} e^{-\eta^2} \quad (11)$$

and therefore the dispersion relation becomes

$$1 - \sum_s \frac{\omega_{ps}^2}{k^2 v_{Th,s}^2} Z' \left(\frac{\omega}{k \cdot v_{Th,s}} \right) = 0 \quad (12)$$

With [8] $Z'(\eta) = -2[1 + \eta Z(\eta)]$. And the expression of Landau damping is given by:

$$1 + \frac{e^2}{\epsilon_0 m_e k^2} \int \frac{k \nabla_v f_0}{\omega - kv} d^3 v = 0 \tag{13}$$

5 STUDY OF ION SOUND WAVES PROPAGATION

5.1 Kinetic theory

The general kinetic dispersion relation for electrostatic waves takes the form

$$1 + \frac{e^2}{\epsilon_0 m_e k} \int_{-\infty}^{\infty} \frac{\partial F_{0e}/\partial u}{\omega - ku} du + \frac{e^2}{\epsilon_0 m_i k} \int_{-\infty}^{\infty} \frac{\partial F_{0i}/\partial u}{\omega - ku} du = 0 \tag{14}$$

And the Landau Damping is given by:

$$1 + \frac{e^2}{\epsilon_0 m_e k^2} \int \frac{k \nabla_v f_0}{\omega - kv} d^3 v = 0 \tag{15}$$

Where

$$F_{0e/i}(u) = \frac{n}{(2\pi T_{e/i}/m_{e/i})^{1/2}} \exp\left(-\frac{m_{e/i} u^2}{2T_{e/i}}\right) \tag{16}$$

The wave with a phase velocity, ω/k , is much less than the electron thermal velocity, but much greater than the ion thermal velocity. We may assume that $\omega \gg k u$ for the ion term. It follows that, to lowest order, this term reduces to $-\omega_{pi}^2/\omega^2$.

Conversely, we may assume that $\omega \ll k u$ for the electron term. Thus, to lowest order we may neglect ω in the velocity space integral. Assuming F_0 to be a Maxwellian with temperature T_e , the electron term reduces to

$$\frac{\omega_{pe}^2 m_e}{k^2 T_e} = \frac{1}{(k \lambda_D)^2} \tag{17}$$

With λ_D is the Debye length [9] and ω_{pe} the electron plasma frequency. Thus, to a first approximation, the dispersion relation can be written

$$1 + \frac{1}{(k \lambda_D)^2} - \frac{\omega_{pi}^2}{\omega^2} = 0 \tag{18}$$

$$\omega^2 = \frac{\omega_{pi}^2 (k \lambda_D)^2}{1 + (k \lambda_D)^2} = \frac{T_e}{m_i} \frac{k^2}{1 + k^2 \lambda_D^2} \tag{19}$$

For $k \lambda_D \ll 1$, we have $\omega = (T_e/m_i)^{1/2} k$, a dispersion relation which is like that of an ordinary sound wave, with the pressure provided by the electrons, and the inertia by the ions. As the wave-length is reduced towards the Debye length, the frequency levels off and approaches the ion plasma frequency.

In the long wave-length limit, we see that the wave phase velocity $(T_e/m_i)^{1/2}$ is indeed much less than the electron thermal velocity [by a factor $(m_e/m_i)^{1/2}$], but that it is only much greater than the ion thermal velocity if the ion temperature, T_i , is much less than the electron temperature, T_e .

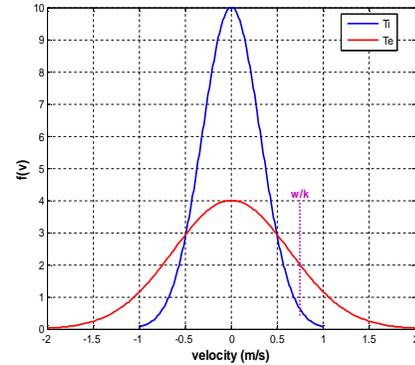


Fig. 3. Ion and electron distribution function $f(v)$ with $T_i \ll T_e$

In fact, if $T_i \ll T_e$ then the wave phase velocity can lie on almost flat portions of the ion and electron distribution functions, as shown in Fig. 2, implying that the wave is subject to very little Landau damping. Indeed, an ion sound wave can only propagate a distance of order its wave-length without being strongly damped provided that T_e is at least five to ten times greater than T_i .

5.2 Fluid theory

Of course, it is possible to obtain the ion sound wave dispersion relation, $\omega^2/k^2 = T^2/m_i$, using fluid theory. The kinetic treatment used here is an improvement on the fluid theory to the extent that no equation of state is assumed, and it makes it clear to us that ion sound waves are subject to strong Landau damping (i.e., they cannot be considered normal modes of the plasma) unless $T_i \ll T_e$.

A. Fluid sound waves

If we perturb and linearize the momentum equation and the continuity equation for field free plasma [10], we get:

$$-\omega \rho_0 \vec{v}_1 = -i \vec{k} \gamma \frac{P_0}{\rho_0} \rho_1 \tag{20}$$

And

$$-i \omega \rho_1 + i \vec{k} \cdot (\rho_0 \vec{v}_1) = 0 \tag{21}$$

Dotting the first equation (20) with \vec{k} and substituting into the second (21), we get:

$$-\omega \rho_1 + \rho_0 \frac{k^2}{\omega \rho_0} \gamma \frac{P_0}{\rho_0} \rho_1 = 0 \tag{22}$$

and the dispersion relation is

$$\omega^2 = k^2 \gamma \frac{P_0}{\rho_0} = k^2 c_s^2 \tag{23}$$

With $v_\phi = v_g = c_s$. This relation is shown on the Fig. 4. as a line which passed by the origin makes into account the proportionality between ω and c_s .

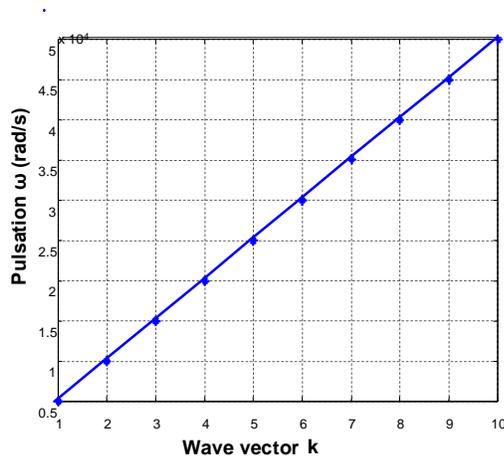


Fig. 4. The dispersion relation $\omega^2 = k^2 c_s^2$

B. Ion sound waves

Even if collisions are unimportant, sound waves, being longitudinal waves, generate density fluctuations which in turn generate electric fields that can provide the necessary restoring force. When ion motion is involved, we know that the waves must be low frequency, so we can use the *plasma approximation*, $n_e \approx n_i \approx n_0$. We are still assuming that there is no magnetic field [9], [10].

$$\frac{\partial}{\partial t}(n_e - n_i) = -n_0 \vec{\nabla} \cdot (\vec{v}_{e1} - \vec{v}_{i1}) \quad (24)$$

Thus if the ion and electron velocities differ, the densities will become different too. Thus the plasma approximation also requires $\vec{v}_{e1} = \vec{v}_{i1}$

$$\omega^2 = k^2 \left(\frac{\gamma_i k_B T_i + \gamma_e k_B T_e}{m_e + m_i} \right) \quad (25)$$

1. It is essentially identical to the result for fluid sound waves even though at a microscopic level there are profound differences. The coupling here is electrostatic non collisional.
2. The electrons move very rapidly, and the distribution may be assumed to be isothermal, $\gamma_e = 1$.
3. The electron mass is negligible compared with the ion mass in the denominator.

However, Vlasov theory (a detailed study of the effect of the particle velocity distributions) shows that the wave is strongly damped unless the electron temperature greatly exceeds the ion temperature. Thus the ion sound speed is determined by the electron temperature and the ion mass.

$$v_{is} = \sqrt{\frac{k_B T_e}{m_i}} \quad (26)$$

We may now consider the electric field necessary to affect the coupling. Poisson's equation is:

$$\vec{k} \cdot \vec{E} = k^2 \phi = \frac{e}{\epsilon_0} (n_i - n_e) \quad (27)$$

Now we allow for small differences between the electron and ion densities. The ion density is given by the continuity equation:

$$n_i = n_0 + \frac{\vec{k} \cdot \vec{v}}{\omega} n_0 \quad (28)$$

while the electrons respond rapidly to the electric field, and so follow the Boltzman relation:

$$n_e = n_0 \exp\left(\frac{e\phi}{k_B T_e}\right) \approx n_0 \left(1 + \frac{e\phi}{k_B T_e}\right) \quad (29)$$

Thus

$$k^2 \phi = \frac{en_0}{\epsilon_0} \left(\frac{\vec{k} \cdot \vec{v}}{\omega} - \frac{e\phi}{k_B T_e} \right) \quad (30)$$

Rearranging, we get:

$$\phi \left(k^2 + \frac{en_0}{\epsilon_0} \frac{e\phi}{k_B T_e} \right) = \frac{en_0}{\epsilon_0} \frac{\vec{k} \cdot \vec{v}}{\omega} \quad (31)$$

We should recognize the second term in the parentheses as $1/\lambda_D^2$. Now we rewrite the momentum equation for the ions, substituting this expression for ϕ in the electric field term

$$(-\vec{E} = -\vec{\nabla} \phi = -i\vec{k} \phi) \quad (32)$$

$$- \omega m_i n_0 \vec{k} \cdot \vec{v}_1 = -en_0 \frac{\epsilon_0}{\epsilon_0} \frac{\vec{k} \cdot \vec{v}_1}{\omega} \left(\frac{1}{1 + k^2 \lambda_D^2} \right) - k^2 \frac{\gamma_i k_B T_i}{m_i} \vec{v}_1$$

$$\omega^2 = k^2 \left(\frac{\gamma_i k_B T_i}{m_i} + \frac{\lambda_D^2 \omega_p^2}{1 + k^2 \lambda_D^2} \right) \quad (33)$$

where ω_{pi} is the ion plasma frequency. The numerator in the second term is:

$$\lambda_D^2 \omega_p^2 = \frac{\epsilon_0 k_B T_e}{en_0} e^2 \frac{n_0}{m_i \epsilon_0} = k_B \frac{T_e}{m_i} \quad (34)$$

Thus the new result is identical to the previous one except for the denominator $1 + k^2 \lambda_D^2$.

Thus the correction is necessary only when $k\lambda_D$ is not small, that is when the wavelength is less than or equal to the Debye length. The full wavelength is within the region where we would expect the plasma approximation to fail. When $k\lambda_D \gg 1$ we find $\omega \approx \omega_p$ and we have oscillations at the ion plasma frequency.

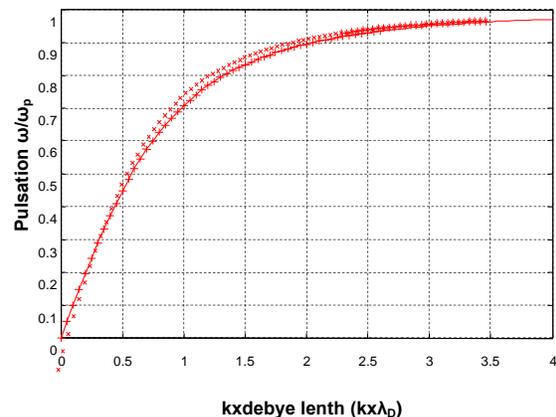


Fig 5: Dispersion relation for ion sound

The wave reduces to plasma oscillations of the ions for which we obtain a graph of the dispersion relation as is shown on the Fig.5. similar to that of the Langmuir wave dispersion relation.

5 CONCLUSION

We have studied the ion-sound waves under the condition of $k \lambda_D \ll 1$, in a hot, isotropic, and unmagnetized plasma modeled with the generalized $f_s(\vec{r}, \vec{v}, t)$ distribution function. We have derived the dispersion relations for the ion-sound waves. It is possible to obtain the ion sound wave dispersion relation, $T_i \ll T_e \cdot \omega^2 / k^2 = T_e / m_i$, using fluid theory. The kinetic treatment used here is an improvement on the fluid theory to the extent that no equation of state is assumed, and it makes it clear to us that ion sound waves are subject to strong Landau damping unless $T_i \ll T_e$. For this condition the electrons are hot and an electrostatic wave in which ions do play a major role is found at lower frequencies. These waves are characterized by phase velocities lying between the thermal velocities of ions and that of the electrons and propagate only when $\omega \gg \omega_p$.

The present work can be extended to study the generation of ion sound waves in the topside low latitude ionosphere at sunrise and sunset.

ACKNOWLEDGMENT

All praises and thanks are for Almighty ALLAH Who is the source of all knowledge and wisdom endowed to mankind and to the Holy prophet Muhammad (peace be upon him) who showed light of knowledge to the humanity as a whole.

REFERENCES

- [1] F. Mottez, "Equations plasmiques pour amateurs de machin-choses compliqués", cour, 11 janvier 2008.
- [2] N.G.Sabri, T.Benouaz, "Study of the Propagation of an electromagnetic wave in a Tokamak Plasma-Wave Plasma Interaction". "PhD Thesis, U. T, 200 pages, 2010.
- [3] M.Moncuquet, "Cours Méthodes numériques de DESS, Laboratoire d'Etudes Spatiales et d'Instrumentation en Astrophysique (LESIA), Paris 2001.
- [4] P.Bertrand, "Description cinétique de la turbulence dans un plasma de Tokamak", Ecole d'Aquitaine onde & matière, Maubuisson 2004.
- [5] E.Sonnendrücker, "Approximation numérique des équations de Vlasov-Maxwell", Notes du cours de M2, 2010.
- [6] R. Fitzpatrick, "Course of the Physics of Plasmas", the University of Texas at Austin, 19-12, 2008.
- [7] Westerhof, "Kinetic theory of plasma waves- homogenous plasma" ; *Transaction of fusion science and technology*, Vol 49, Feb. 2006.
- [8] J. Jiménez-Mier, "An approximation to the plasma dispersion function", *Journal of quantitative Spectroscopy & Radiative Transfer* 70 (2001) 273-284.
- [9] I.Weston, M.Stracey, *Fusion plasma physics*, WILEY-VCH, pp. 133-134, 2005, ISBN-13:978-3-527-40586-2. (Book style)
- [10] S.M.Lea, "Plasma Waves", course; January 2007, unpublished.

A Proposed Solution for Sorting Algorithms Problems by Comparison Network Model of Computation.

Mr. Rajeev Singh, Mr. Ashish Kumar Tripathi, Mr. Saurabh Upadhyay, Mr.Sachin Kumar Dhar Dwivedi

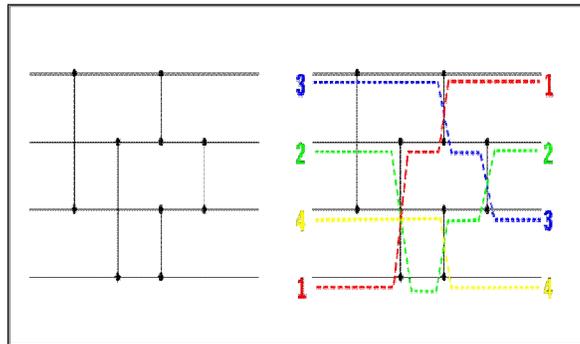
Abstract:-In this paper we have proposed a new solution for sorting algorithms. In the beginning of the sorting algorithm for serial computers (Random access machines, or RAM'S) that allow only one operation to be executed at a time. We have investigated sorting algorithm based on a comparison network model of computation, in which many comparison operation can be performed simultaneously.

Index Terms

Sorting algorithms, comparison network, sorting network, the zero one principle, bitonic sorting network

1 Introduction

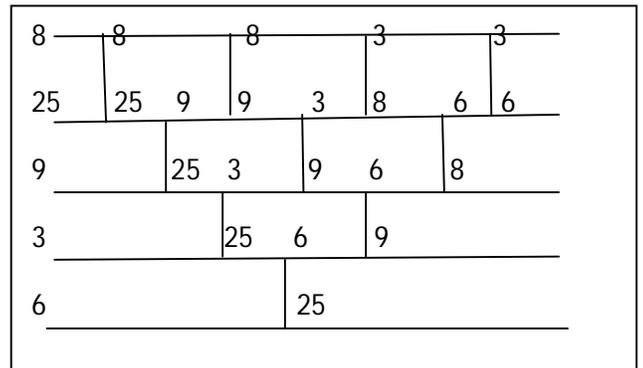
There are many algorithms for solving sorting algorithms (networks).A sorting network is an abstract mathematical model of a network of wires and comparator modules that is used to sort a sequence of numbers. Each comparator connects two wires and sorts the values by outputting the smaller value to one wire and the large to the other. A sorting network consists of two items comparators and wires .each wires carries with its values and each comparator takes two wires as input and output. This independence of comparison sequences is useful for parallel execution of the algorithms. Despite the simplicity of the model, sorting network theory is surprisingly deep and complex.



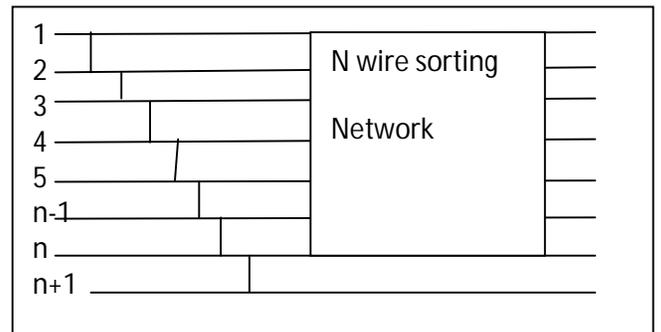
A **sorting algorithm** is an algorithm that puts elements of a list in a certain order. The most-used orders are numerical order. Efficient sorting is important for optimizing the use of other algorithms that require sorted lists to work correctly; it is also often useful for data and for producing human-readable output. More formally, the output must satisfy two conditions:

- 1.1 The output is in no decreasing order (each element is no smaller than the previous element according to the desired total order);

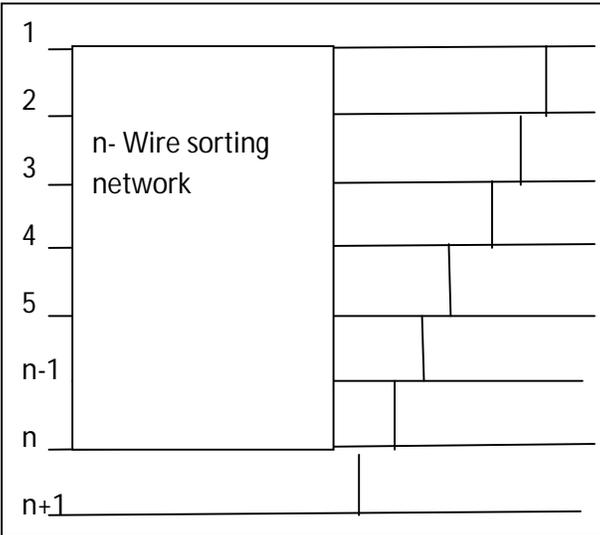
- 1.2 The output is a permutation, or reordering, of the input. For example of bubble sort 8, 25,9,3,6



We can easily construct a network of any size recursively using the principles of insertion and selection. Assuming we have a sorting network of size n , we can construct a network of size $n + 1$ by "inserting" an additional number into the already sorted subnet . We can also accomplish the same thing by first "selecting" the lowest value from the inputs and then sort the remaining values recursively (using the principle behind bubble sort).

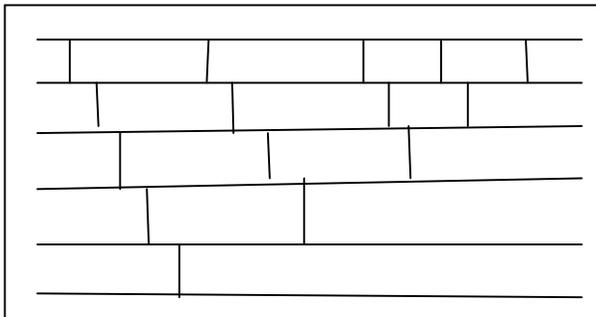


A Sorting Network constructed recursively that first sinks the largest value of the bottom and then sorts the remaining wires. Based on bubble sort.

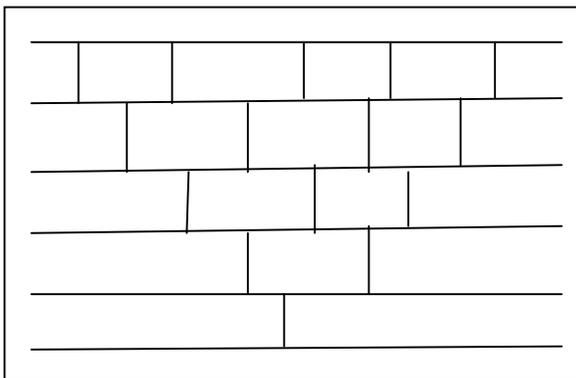


A Sorting Network constructed recursively that first sort the first n wires and then sorts the remaining wires. Based on Insertion sort.

The structure of these two sorting networks is very similar. A construction of the two different variants, which collapses together comparators that can be performed simultaneously shows that, in fact, they are identical



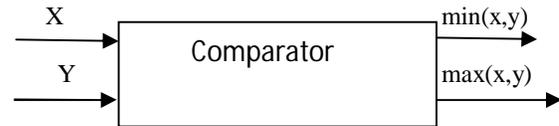
Bubble sorting network



Insertion Sorting Network.

2 Comparator network

A sorting network consists of two items: comparators and wires. Each wire carries with it a value, and each comparator takes two wires as input and output. When two values enter a comparator, the comparator emits the lower value from the top wire, and the higher value from the bottom wire. A network of wires and comparators that will correctly sort all possible inputs into ascending order is called a sorting network.



A Comparator is a mapping $\{I, j\}$
 $A^n \rightarrow A^n$ $I, j \in \{0, \dots, n-1\}$
 $\min(X, Y)$

With

$\{i, j\}$ (a) $I = \min(a_i, a_j)$

$\max(X, Y)$

$\{i, j\}$ (a) $j = \max(a_i, a_j)$,

$\{I, j\}$ (a) $k = a_k$ for all k with $k \neq I, k \neq j$

For all $a \in A^n$

3 Objectives

To present that is one after all operations in a comparison network may occur at the same time or “in parallel” We investigate sorting algorithm based on a comparison network model of computation, in which many comparison operations can be performed simultaneously.

4 The zero-one principle

The zero one principle says that if a sorting networks work correctly when each input is drawn from the set $\{0,1\}$, then it works correctly an arbitrary input numbers. The numbers can be integers, reals or in general, any set of values from any linearly ordered set. as we construct sorting network and other comparison networks the zero one principle will allow us to focus on their operation for input sequence consisting solely of 0’s and 1’s. it is easy to prove the validity of some sorting networks (like the insertion sort, bubble sort), it is not always so easy. There are $n!$ permutations of numbers in an n -wire network, and to test all of them would take a significant amount of time, especially when n is large. The number of test cases can be reduced significantly, to 2^n , using the so called zero-one principle. the zero one principle states that a sorting networks is valid if it can sort all 2^n sequence of 0’s and 1’s.

If a comparison network transforms the input sequence $a = \{a_1, a_2, \dots, a_n\}$ into the output sequence $b = \{b_1, b_2, \dots, b_n\}$ then for any monotonically increasing function f , the network transforms the input sequence $f(a) = \{f(a_1), f(a_2), \dots, f(a_n)\}$ into the output sequence $f(b) = \{f(b_1), f(b_2), \dots, f(b_n)\}$.

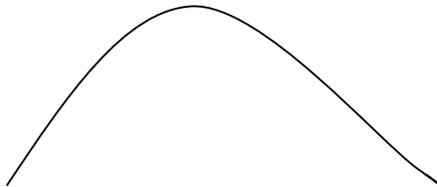
5 Bitonic sorters

The first steps in our construction of an efficient sorting network is to construct a comparison network that can sort any bitonic

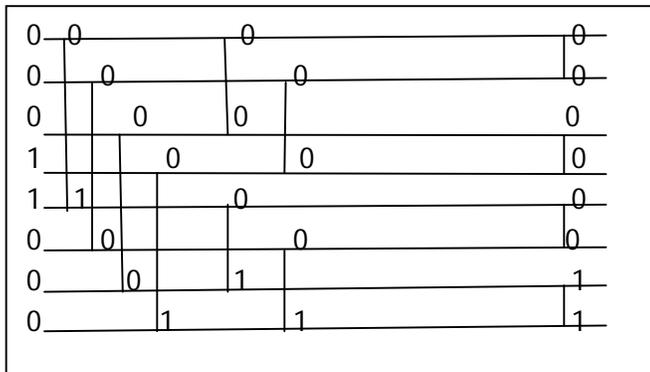
sequence, a sequence that monotonically increase and monotonically decrease or can be circularly shifted to become monotonically increasing and then monotonically decreasing. For example, the sequences <1,4,6,8,3,2>, <6,9,4,2,3,5> and <9,8,3,2,4,6> are all bitonic. The bitonic sorter that we shall construct is a comparison network that sorts bitonic sequence of 0's and 1's. for example 0000-----111-----0000.....



For example 1, 3,9,22,56,98,95,2511,4,2



Eg 00011000



If no of 0's > no of 1's then in upper half all zero's and in lower half a bitonic sequence, if no of 1's > no of 0's then lower half completely 1's and upper half a bitonic sequence, if no of 1's = no of 0's upper half all zero & lower half all 1's. no of comparator used in bitonic sorter = n log n. Complexity of bitonic sequence = O(log₂n), where n = no of I/P.

A bitonic sorter is composed of several stages, each of which is called half cleaner, each half cleaner is a comparison network of depth 1 in which input line I is compared with line i+n/2 for i=1,2,...,n/2. sort the sequence 18,22,56,98,102,88,76,24 using bitonic is complexity log₂n

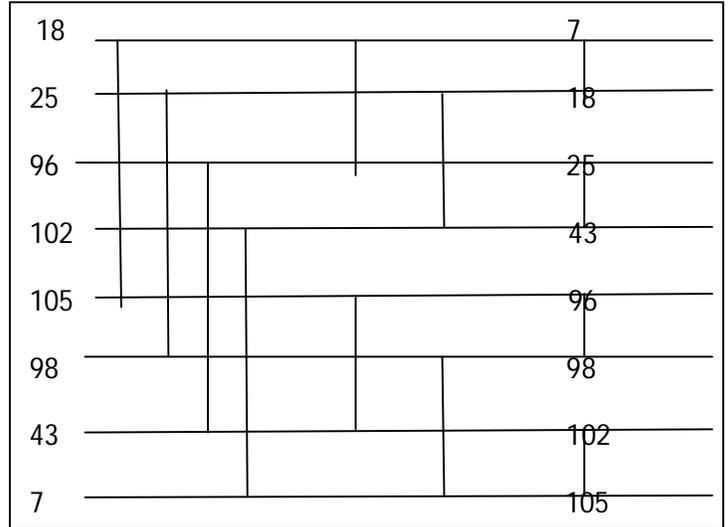
Merging network:(V)-our sorting network will be constructed from merging will be constructed from merging networks, which are networks that can merge two sorted input sequence into one sorted input sequence .we modify bitonic sorter[n] to create the merging network merger [n].

For example, given the sorted zero one sequence
X=00000000111

Y=00000001111, we reverse y to get yR 11110000000

Concatenating X and yR yields. Thus to merge the two input sequence X and Y, it suffices to perform a bitonic sort on x concatenated with yR.

For example s1: 18, 25, 96, 102, and 105,
s2:7, 43, 98 merging of two sorted sequence using bitonic method.



6 Analysis of sorting network

We now have all the necessary tools to construct a network that can sort any input sequence. The sorting network sorter[n] uses the merging network to implement a parallel version of merge sort. We can analyze the depth of the sorting network recursively. The depth d(n) of sorter [n] is the depth log₂n of merger [n] consequently, the depth of sorter[n] is given by the recurrence.
D[n] = 0 if n=1
D (n/2) + logn if n=2k and k>=1

We can sort n numbers in parallel in (log₂ n) time. Counting sort cannot be implemented on a comparison network.

7 CONCLUSION

Thus an algorithm such as counting sort cannot be implemented on a comparison network. Second the RAM model in which operations occur serially that is one after another operations in a comparison network may occur at the same time or in parallel". As we shall see, this characteristic allows the construction of comparison network that sort n values in sub linear time.

8 References

1Improved Algorithms and analysis for secretary problems and generalizations. In proceeding of the 36th annual Symposium on foundation of computer science.

- 2 Mohammad Akra and Louay Bazzi: on the solution of linear recurrence equations.computational optimization and applications.
3. G.M Adel's on-vel'skii and E.M Landis.An algorithm for the organization of information .soviet mathematics doklady.
4. Eric Bach and Jeffery Shallit.algorithmic number theory-volume-I: Efficient algorithms, The MIT press, 1996
- 5 R.Bayer and E.M McCreight.Organization and maintenance of large ordered indexes.Acta Information 1(3):173-189, 1972.
6. Pierre Beauchemin, Gilles Brassard, and Claude Creapeau and Carl Pomerance.The generation of random numbers that is probably prime.
7. Yinyu Ye.Interior Point Algorithms: Theory and analysis, john wiley \$Sons, 1997
8. JeanVuillemin.Adata structure for manipulating priority queues.coomunication of the ACM.
9. Mark Allen Weise.Data Structures and Problems Solving Using Java.
10. C.A.R .Hoare .Merge sort. Computer Journal, 5(1): 10-15, 1962
11. O. Angel, A.E. Holroyd, D. Romik, B. Virag, *Random Sorting Network* Adv. in Math., 215(2):839–868, 2007.
12. K.E.Batcher, sorting network and their application Proceedings of the AFIPS Spring Joint Computer Conference 32, 307–314 (1968).
13. D.E. Knuth the, Volume 3: *Sorting and Searching*, Third Edition. Addison-Wesley, 1997.ISBN 0-201-89685-0 Section 5.3.4: Networks for Sorting, pp. 219–247.
- 14.Ajtai, M.; Komlós, J.; Szemerédi, E. (1983), "An $O(n \log n)$ sorting network", *Proceedings of the 15th Annual ACM Symposium on Theory* .
15. M. S. Paterson, *Improved sorting networks with $O(\log N)$ depth*, *Algorithmica* 5 (1990), no. 1, pp. 75–92, doi:10.1007/BF01840378.
16. M. Mitchell, *an Introduction to Genetic Algorithms*, the MIT Press, 1998.ISBN-0-262-63185-7. Chapter 1: Overview, pp. 21–27

Author:-

- 1 Mr. Rajeev Singh is currently working in Dept. of CS, Centre for Management Technology Greater Noida, U.P, India
E-mail:-rajeevvisodiamca@gmail.com
- 2 Mr. Ashish Kumar Tripathi is currently working in Dept. of CS, Centre for Management Technology Greater Noida, U.P, India
E-mail:-ashish.msa@gmail.com
- 3 Mr.Saurabh Upadhyay is currently working in Dept. of CS, S.M.S Varanasi, U.P, INDIA
E-mail:- saurabhupad@gmail.com
- 4Mr.Sachin Kumar Dhar Dwivedi, Assistant Engineer National Institute of Electronics and Information Technology, Aizawl, India.
E-mail:-sachin8071@gmail.com

“Merging of independent MANETs- RCF process model: Check the performance Delay and throughput of server when IP Address conflict Occurs”

Chand Mukesh*, H.L.Mandoria#

*Research Scholar, Graphic Era University, Dehradun, India

Prof. CoE&T, PantNagar University, PantNagar, India

mukesh.geu@gmail.com

Abstract

As the number of nodes increases the complexity of MANET [1, 11] increases in various issues. For this reason various approaches has been produced to reduce the complexity such as cluster head technique and dominating set based gateway technique. Another issue is distribution of IP in MANET. There are various approaches is given to assign the IP address but they are not much effective and each and every approach has its limitation. In this paper a conflict free process model is proposed to solve the configuration problem where two or more than two MANET merge and remove the limitation from the above-mentioned approaches .And also the performance of server due IP address conflict is tested with the help of a scenario to solve IP address configuration [2, 4, 8, 13, and 16] problem in mobile adhoc network. Simulation in this thesis shows that duplicates address in a MANET affects the system in various ways.

Key-Words: Mobile Adhoc Networks, IP Address, Private Addressing Scheme

1. Introduction:

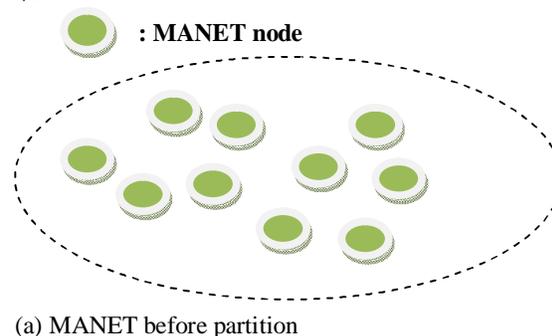
Mobile adhoc networks are infrastructureless self-organizing wireless networks. Each node can be mobile and has routing capabilities to be able to forward packets on behalf of other nodes Adhoc networks are typically composed of homogeneous nodes that communicate over wireless links without any central control. Adhoc wireless networks inherit the traditional problem of wireless and mobile communication, such as bandwidth optimization, power control and transmission quality enhancement .In addition topology is highly dynamic & random & very hard to predict. Physical security is limited. Mobile Ad-hoc Network serves as a temporary wireless network in which node changes its IP address with the help of an intelligent auto-configuration protocol [3, 14, 18]. The main role of IP address auto-configuration protocol is to manage the address space .The protocol must be able to allocate a unique network address [4] to un-configured node.

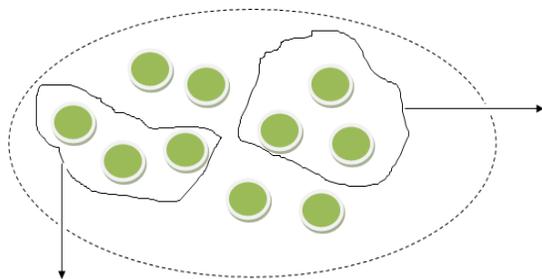
2. Related Work:

There are several scenarios in which a mobile node will change its IP address:

i) Partitions of a network in MANET[7,15]

If some mobile nodes in the MANET move out of the transmission range of the other nodes, the network becomes partitioned. Because these nodes may not be aware of the partition, they may still use the previous allocated addresses. If IP address of a node in one partition is allocated to the new node in the other partition, address conflict occurs when these two partitions become connected. If some mobile nodes in the MANET move out of the transmission range of the other nodes, the network becomes partitioned as in figure 1. Because these nodes may not be aware of the partition, they may still use the previous allocated addresses. If IP address of a node in one partition is allocated to the new node in the other partition, address conflict occurs when these two partitions become connected. Partition of adhoc network is demonstrated as in Fig. 1, 2 and 3.





(b) When some of the nodes start moving

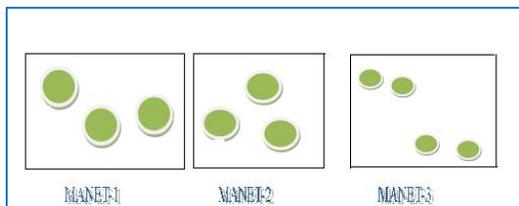


Fig. 2: After partition the network is partitioned into three different MANET

Due to this partition node in each partitioned MANET carry the same IP address as in before partition. The problem arises when these MANETs are merged than the possibility of collision may occur because of address conflict.

ii) Merger of two independent Mobile networks [7, 15]

The second scenario is that two independent configured MANETs are merged as in Fig. 4. And the MANET before merging is shown in Fig. 1. Because these two networks are auto configured separately, there may be some duplicate addresses in both networks, such as node A in MANET1 and node B in MANET2. Thus one needs to change its addresses due to the merger.

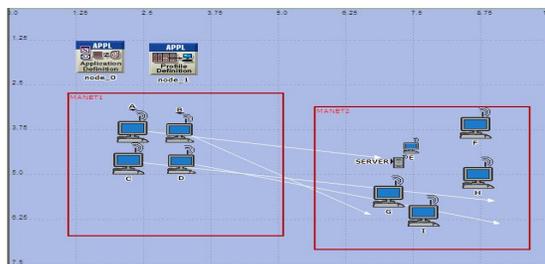


Fig 1: Merger of two networks

3. Problem Formulation

To overcome IP address configuration problem A Frame Work is proposed in the previous paper

[19].Proposed framework is divided in to three phases, each phase comprising its own level of complexity, and the aim is to achieve the optimum solution from the combined efforts of each phase. The lower phase has various level of complexity such as merger of two networks and partition of one single network into multiple networks, mobility of each node which causes the frequent disconnection of the node. In this paper we are taking the first phase of addressing of different scenarios. In this when two independent networks are merging then this leads the high degree of probability that some nodes are using same IPs, and during merging of two networks it will create the problem of conflict, which must be resolved before merging of the network. To resolve configuration problem a Proposed algorithm is applied to the MANETs (this may also be applied when number of MANETs are increases and hence solve the problem of scalability).

Process Model for IP configuration: This process works when a mobile node(s) finds another node in the network. It sends a signal to initiate the request by using a hello Message and waits until a response comes. The node in this Reconfiguration process model (Fig. 2) works as a network head.

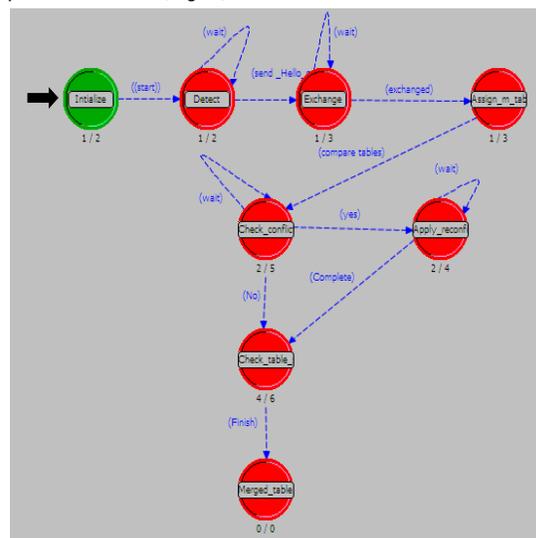


Fig. 2:RCF Process Model 1

Initialize: when a process starts. (Start)

Detect: When a node(s) finds (send_Hello_mesg) another nodes in the network and wait for the response

Exchange: At this state nodes Exchange their routing table and update the latest entries. (exch_routing_tab).After table enteries are exchanged table entries are updated at this state and marked as a new table.

Assign_m_table: AT this state the size of the table is checked and if the size of the node1 is greater than the size of node2 then the node1 will be the new head for the merged network.

Merged_table: This state results a merged table or complete table having all the entries in the MANET.

Check_Conflict: In this state table entries are checked and if conflict occurs the reconfiguration process is applied.

Apply_Reconfig_Procedure: Reconfiguration process is applied at this stage.

Check_table_size: In this state if there is no conflict after merging then the size of each table (MANETs) is checked .The network head will be chosen from the largest table size MANET.

St_8: Check table size

Set: Assign a random number to blank entries.

The background details and functions of each node in this process model is explained in the forthcoming paper. For ease of computation we are assuming that in the Scenario there is one network head that is having the central authority for managing the IP addressing, a Process Model is proposed to solve the various cases where IP addresses conflict occurs .The IP address table Tab. 1 describe IP addresses for this simulation. Node from A TO D is assigned to MANET1 and node from E TO I is assigned to MANET2.

Name Of Node	IP Address	Host number
A	192.168.1.1	1
B	192.168.1.2	2
C	192.168.1.3	3
D	192.168.1.4	4
E	192.168.1.5	5
F	192.168.1.6	6
G	192.168.1.7	7
H	192.168.1.8	8
I	192.168.1.9	9

The Remaining Addresses may be assigned to the duplicate node/new node when an address conflict occurs

Tab. 1: IP Address table

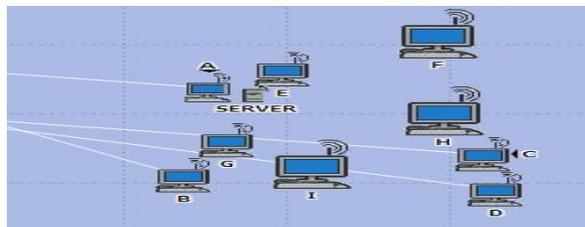


Fig. 4: After merging two MANETs

The table Tab. 2 described all the parameters used

Result Analysis: In this paper two nodes (1, 5) are selected to check the performance of the server. The performance is checked by using the private IP address as follows:

Case 1: When the node A is assigned an IP address 192.168.1.1 and the E is assigned an IP address 192.168.1.5, the load on the server is simulated .In this simulation observation are taken in terms of request (Fig. 5) on the server, Total traffic (Fig. 6) delay. Load, packet dropped, throughput and the number of packet dropped (fig. 7) during simulation as shown in Fig. 5, 6, 7 & 8 and simulation data is shown in table 3.

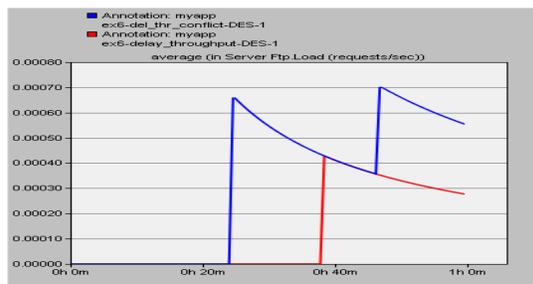


Fig. 5: load (request/sec)

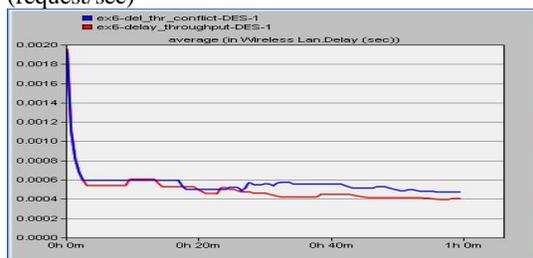
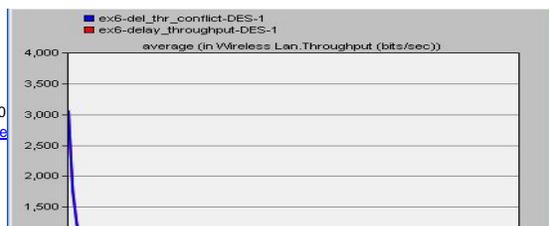


Fig 6: Delay (sec)



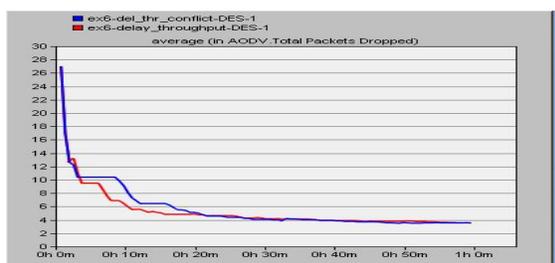


Fig 8: Packet dropped (packets/sec)

Case 2: When the node A is assigned an IP address 192.168.1.1 and the E is assigned an IP address 192.168.1.1, the load, throughput, delay, packet dropped on the server is simulated as in Fig. 5, 6, 7 and 8. In this simulation observation are taken in terms of request on the server. In this case an IP address conflict occurs because of duplicate address. The server treats the request (coming from A after moving to server and from E) as the same and response both the nodes at once. By comparing dropped packets we have found that in this case the rate of packet dropped is very high and hence information is lost in this case. Throughput, delay and Load also increases when the conflict occurs which results in performance degradation. The simulation data are shown Table 3.

Conclusion & Further Discussion:

In this paper a process model is designed to solve the configuration problem when two or more than two MANETs are merge to assign the conflict free IP address in the network. The information loss and delay in response is tested when IP address conflict occurs with the help of a scenario. Simulation result shows that Load, Throughput, Packet dropped are also affected when address conflict occurs. Further discussion may include the designing of a node model and add the above process side by side as an alternative to DHCP. The aim of the research is to implement all the phases of the framework for IP Address configuration in MANET given in previous paper.

References:

- [1] S. Corson University of Maryland, J. Macker Naval Research Laboratory January 1999 <https://www2.ietf.org/rfc/rfc2501.txt>
- [2] Kilian Weniger and Martina Zitterbart, University Karlsruhe (TH), "Address Auto configuration in Mobile Adhoc Network: Current Approaches and Future Directions", IEEE Network, July/August 2004.
- [3] M. Nazeeruddin, G. Parr, Member, IEEE and B. Scotney, Member, IEEE Fault-tolerant Dynamic Host Auto-configuration Protocol for Heterogeneous MANETs ["www.eurasip.org/Proceedings/Ext/IST05/papers/230.pdf"](http://www.eurasip.org/Proceedings/Ext/IST05/papers/230.pdf)
- [4] Taghiloo, M.; Dehghan, M.; Taghiloo, J.; Fazio, M.; Comput. Eng. Dept., Amirkabir Univ. of Technol., Tehran "New Approach for Address Auto-Configuration in MANET Based on Virtual Address Space Mapping (VASM)", *ICTTA, 2008*, <http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=4520396>
- [5] Christian Tschudin and Richard Gold "Lightweight Underlay Network Ad-Hoc Routing", April 2002. <http://www.docs.uu.se/selnet/lunar/>
- [7] Hongbo Zhou and Matt W. Mutka, Lionel M. Ni. "IP Address Handoff in the MANET" IEEE INFOCOM 2004.
- [8] Dongkeun Lee, Jaepil Yoo, Keecheon Kim and Kyunglim Kang "IPv6 Stateless Address Auto-configuration in Mobile Ad-Hoc Network", <http://www.springerlink.com/content/73m5844032323822/>
- [9] D. Senie. RFC-3235 - Network Address Translator (NAT)-Friendly Application Design Guidelines, January 2002.
- [10] Christophe Jelger and Christian Tschudin Technical Report CS-2006-001 University of Basel June 12, 2006. "Private Address Maps (The Disappearing MANET Autoconf Problem)". http://persons.unik.no/paalee/referencing_publications/ref-autoc-jelger-techrep06.pdf.
- [11] IETF MANET Working Group. MANET Charter, 2000. <http://www.ietf.org/html.charters/manet-charter.html>.
- [12] M.Q.Rafiq, S.R Chauduri, S.Kumar, S.Husain "Deciding the Cluster Size on Behalf of Co- channel Interference Ratio for the Omni Directional Antenna in Mobile Computing" ISTE day conference New Delhi Chapter 31st May & 1st June, 08.
- [13] C. Bernardos and M. Calderon. Internet Draft - Survey of IP address auto configuration mechanisms for MANETs (expired), July 2005.
- [14] P. Engelstad, D. V. Thanh, and G. Egeland. Name Resolution in On-Demand MANETs and over External IP Networks. In *Proceedings of IEEE ICC'03*, May 2003. Anchorage, Alaska.
- [15] P. Krishna, M. Chatterjee, N.H Vaidya, & D.K Pardhan, "A Cluster based Approach for routing in adhoc networks", Proceeding of 2nd USENIX Symposium on Mobile and Location Independent Computing. P.1 1995
- [16] N. Vaidya. "Weak Duplicate Address Detection in Mobile Ad hoc Networks," ACM MobiHoc 2002, Lausanne, Switzerland, June 2002, p 206-216.
- [17] Shenglan Hu, Chris J. Mitchell, "Improving IP address auto configuration security in MANETs using trust modelling,"

Mobile Ad-hoc and Sensor Networks - First International Conference, MSN 2005, p 83-92.

[18] K. Weniger. "Pacman: Passive auto configuration for mobile ad hoc networks," IEEE Journal on Selected Areas in Communications, Special Issue 'Wireless Ad Hoc Networks', March 2005, p 507 - 519.

[19] M.Chand "A PROPOSED ALGORITHM AND A MODEL FOR IP ADDRESS CONFIGURATION IN MANET", ICACC – 2011 (International Conference on Advances in Computing and Communication) NIT Hamirpur, India.

Time (sec)	Packet Dropped(packet/sec)		Delay(sec)		Load(bits/sec)		Throughput(bits/sec)				
	Conflict	Unique	Conflict	Unique	Conflict	Unique	Conflict	Unique			
0.0					0.010		0.0011	11202	11149	12022	11935
36.0	27	27			0.0029		0.0029	2557	2752	3928	4200
72.0	7	7			0.0026		0.0025	597	578	867	783
108.0	4	4			0.0025		0.0025	698	697	994	970
144.0	11	15			0.0047		0.0037	2560	2679	3960	4888

Table 3: Simulation data

Table 2: Simulation Parameters

S.NO	Name	Description	Value			
1.	Simulation Setup	Mobility Configuration Used to define Mobility Profile for the mobile nodes	Speed(m/s)	10(const)		
			Start Time(s)	10(const)		
			Pause time(s)	100(const)		
	Application Definition	Used to define different Application used in this scenario for all participating nodes	Type Of application	FTP	IRT(s)	10
					File Size(Bytes)	5000
					Typeof Service	Delay, throughput
	Profile Definition	Used to specify the common profile on different nodes in the network	Protocol Used	AO DV	Active Route Time Out(s)	3
					Hello Interval(s)	Min(-1),Max(1.1)
					Network Dia.	35
					Addressing Mode	IPV4
2.	Simulation Parameters	Transmit Power(W)	0.005			
		Packet Reception Power Threshold(W)	-95			
		Simulation Time(hrs)	1			
		No. Of Nodes	9			
		Environment Size	1000*1000(meters)			
		Traffic Type	FTP			

Symbian phone forensics- An agent based approach

Deepa Krishnan, Satheesh Kumar.S, A.Arokiraj Jovith

Abstract— Smart phones like the older mobile phones are fast becoming a life style choice. These sleek devices with large amount of personal information in them make smart phone forensics, a key component in any criminal investigation. The paper presents a contrast between hardware and software approaches and highlights the key advantage of the software approach i.e. the speed at which actionable data can be made available with less technical knowhow. Moreover, the proposed plug-in based agent development provides an extensible framework to handle customizations that will matchup with each unique nuance of phone platform and model. The paper summarizes the finding by unveiling a prototype module development, using platform SDK and on-phone agent. The main focus is the simplicity and extensibility of proposed approach but at the same time the paper does warn about the possible impact to device memory and contrasts with other alternatives.

Index Terms— Cyber forensics, law enforcement, mobile computing, security.

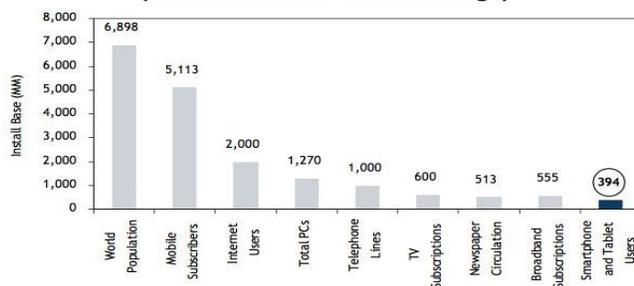
1 INTRODUCTION

THESE is no device that has changed lives and has seen world wide adoption like our humble mobile phone. With the advent of smart phones and its integration with web and social networking we are at the cusp of another radical change. We are in the era where phones have horsepower equivalent to a PC, were phone based news reporting has brought about revolution and downfall corrupt of a regimes. These powerful devices in the wrong hands will be equally disruptive. In this changing background forensic analysis of phones has become even more challenging all when our law enforcement agencies need to process a wide range of handsets quickly and get all available information to the investigating officer.

Today smart phones have almost all the features of a laptop or a notebook computer. Analysis of such devices is a major agenda before the forensics community. The law enforcement agencies require sophisticated software as well as hardware tools for the proper analysis of digital evidence to bring the culprit before the court of law.

If you look at figure 1[9] it shows the current penetrations of mobile phones in relation to world population and how mobile phones usage stacks up in comparison with other technologies. As smart phones replace the current generation of phones we are looking at a massive redefinition of current process.

**Tablet and Smartphone Users vs. Other Markets
 (The Potential For Growth Is Huge)**



Source: RBC Capital Markets

Fig 1 World mobile phone penetration and potential growth of smart phones

The capabilities and features of each handset determine what information could potentially be retrieved from each device. This is easier to understand if we look at this from the perspective of what particular tasks handset could perform. For example, older phones have limited memory so what we can expect to get is limited to data usually form the SIM card. Modern phones on the other hand have huge internal storage, which can further be extended by external memory card. Apart from a camera, many of the modern phones come equipped with GPS, compass, humidity sensor, proximity sensor, gyroscope and much more. The table 1 shows a comparison between different categories of phones.

**TABLE 1
 COMMON ATTRIBUTES OF SMART PHONE**

	Basic	Advanced	Smart
OS	Proprietary	Proprietary	Android,iOS, RIMOS, Palm OS, Symbian, Windows Phone7
Address Book	Basic Book	Phone Address Book with possible Calendar	Elaborate address book including special apps from app store.
Apps	Non existent	Basic pre- build App	Wide range of app selection from app store
eMail & Chat	None	SMS Chat	Wide range of chat & eMail app from the app store
Web	None	None or WAP Gateway	HTTP
Wireless	None	IrDA or Bluetooth	IrDA, Bluetooth, WIFI

Over the past few months the landscape of smart phone market share has drastically changed with Android and iOS making huge gains at the expense and in some cases downfall of the competition.

Before getting into the process details it is appropriate to look at the basic information present in smart phones. These can include but not limited to:

1. Handset Setting (language, date & time, GPRS,

- WAP, internet etc)
- 2. Phone Book
- 3. Call Logs (incoming, missed and outgoing)
- 4. SMS Messages
- 5. Tasks
- 6. Calendar Events
- 7. Stored Files (pictures, music, video, audio recording etc)

2 FORENSIC PROCESS

While dealing with a digital device the method used to acquire data must have little impact on device memory as possible. Impact if any should be well understood and documented. This is important to ensure that integrity of the acquired data and also to allow for a 3rd party verification if it is required at any stage.

Forensics on a cell phone is considerably different from a personal computer. Even though the number of platform we have to deal with is reducing there is still a wide selection of proprietary OS along with a short product release cycle. Hence we will always be dealing with a moving target even within a single platform. Methodology and sequence in which the phone is handled and data collected is critical [4]. Turning off the phone has the potential to alter its memory or data on the phone, but leaving the phone ON raises the possibility of new information arriving over the network and affecting the integrity. Ideally the phone should be placed in a radio shield environment and it should only be switched OFF if that's not possible. On the same lines removal of SIM card or battery from some phones could modify the contents of phone memory. We recognize the complexity of this process and are developing an application based on the plug-in model, which allows for extensive customization based on the phone platform, model and version. The application will identify the expected steps and walk the user through it.

The following section describes some of the things a crime scene technician should consider as he/she goes through the evidence. The application helps the technician with proper reminders and logging the finding.

2.1 Keeping Track of Yourself

General guidelines for forensics require that, investigators cannot change the digital data content of the device being analyzed. Moreover, an audit trail of the analysis and investigation process should be maintained at all times in such a way that it can be verified by multiple sources. Each step should be accurately documented, so that there is enough information for the process to be reviewed by independent third-party. Finally, the person in charge of the investigation should maintain compliance with the governing laws.

Forensic method used should minimize changes on the device, be able to retrieve the full set of data, and finally minimize user interaction with the device itself. Ideally, the full memory content of a generic embedded device should be collected, to preserve the full inner state and obtain a forensically sound acquisition.

It is also recommended to keep track of approach and progress, by means of an external recording device (e.g. camera) that will maintain visual breadcrumbs.

2.2 Background Data

During the intake and processing of the phone evidence the crime scene investigator from law enforcement inputs a bunch of contextual information. This includes but not limited to where the evidence was found, the crime file details, the technician name etc. Capturing this kind of context information that can be kept with the analysis report is the first step of the forensic extraction tool.

2.3 External Forensic Data Source

There is information that can be gleaned from outside vis-à-vis the network can be as important as what is in the phone. For e.g. in the GSM network environment, a great deal of information might be recovered from the service provider. The set of information, which can be successfully collected with this method, is related to the SIM data set, such as SMS, MMS, list of last called numbers, and the location of the subscriber. Clearly, information such as photos, videos, phone book, web browser logs, audio recordings, or user's notes cannot be gathered in such a way.

If external forensic data can be gathered then the request for the information from the service provider or notes regarding this are recorded by the technician.

2.4 Physical Data Extraction

Physical acquisition implies a bit-by-bit copy of the entire physical store. Physical acquisition [5] has its advantages since it allows deleted files and data remnants in unallocated memory to be analyzed. Once a bit-by-bit copy is made the extracted image need to be parsed and decoded manually. Logical extraction of data implies copying data in logical file system partition through OS framework calls. This is a logical view not raw memory view, which has its disadvantages.

There are not many effective tools available to take an effective physical image and parse it to something meaningful; most forensic tools for cell phones and SIMs acquire more technical knowhow and training. At the minimum the technician should know how to hook up to diagnosis or debugging ports like JTAG or at the extreme level may require de-soldering the flash chip and connecting to the reader. NOTE: De-soldering the flash chip is the most invasive method for the equipment so may not be the right approach in all cases. But this is ideal when the phone is damaged. If physical hardware based extraction is deemed useful the technical records those though-

ts in the report log.

2.5 Mobile Phone Communication Interface

Different interfaces [7], [12] like IrDA, Bluetooth or serial cable can be used to acquire logical content. Extracting data using serial cable is the recommended option; wireless options should be used only after understanding the potential forensic issues. E.g. Bluetooth requires the wireless antenna to be switched ON and requires key entries on the handset so that it is paired with the forensic workstation and a good connection is setup all of this generates integrity concerns.

2.6 Logical Data Extraction

This is the heart of the process and relies on multiple protocol and communication methods some of the things used are AT Commands, SyncML, FBUS, MBUS, OBEX, IrMC APDU. As you will see in further sections the phone OS platform SDK provides powerful options to extract data. Because each phone has its own unique approach; the plug-in model provides an extensible option to pick and choose what works best for the phone platform and model.

As you choose the phone model the correct plug-in that will do the job is called and used to extract the information. The extracted information and the final report is then run through a hashing algorithm (MD5) before saving it, to prevent tampering.

3 UNDERSTANDING SYMBIAN PLATFORM FOR PHONE FORENSICS

In the remainder of this paper we will dive deeper into developing the proposed module by choosing one of the smart phone platforms i.e. Symbian S60. We will start the discussion with an overview of the Symbian operating system and lay the groundwork of what we are dealing with. In sections following that describes different methods employed in retrieving data.

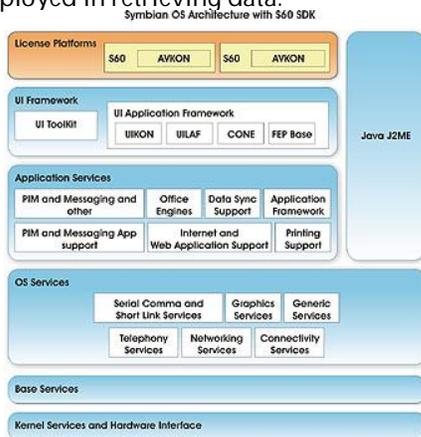


Fig 3 Symbian OS Architecture

3.1 Symbian OS Architecture

As can be seen in Figure 4, the architecture is modular, where operating system functionalities are provided in separate building blocks, and not in one monolithic unit. Being a single purpose phone OS Symbian [2] is single user with multi-tasking capability, being able to switch CPU time between multiple threads, giving the user of the mobile phone the impression that multiple applications are running at the same time.

The core OS is formed by a microkernel built as a personality layer on top of a real-time (RTOS) nanokernel. This block is responsible for primitives such as fast synchronization, timers, initial interrupts dispatching, and thread scheduling. Generally speaking, Symbian OS is intended to run on open, small, battery-powered portable computers, which are modern advanced state-of-the-art mobile phones.

3.2 Symbian File System

On a Symbian smart phone, the file system [5],[13] can be accessed by means of the file server component, also referred to as F32, which manages every file device. It provides services to access the files; directories and drives on those file mapped devices. The file server uses the client/server framework, by receiving and processing file-related requests from multiple clients. Moreover, it is able to deal with different file system formats, such as the FAT format used for removable disks, by using components that are plugged into the file server itself. In addition, it supports a maximum of 26 drives, each identified in DOS-like convention by a different drive letter, in the range A: - Z.

The main ROM drive, where the firmware resides, is always designated as "Z:". This drive holds system executables and data, which are referred to as XIP (eXecutable In Place) because they are directly launched without being loaded into RAM. Besides this, the firmware, or ROM image, is usually programmed into Flash memory, known also as EEPROM, the nonvolatile memory that can be programmed and erased electronically.

The C: drive is always designated as the main user data drive, which can be mapped onto the same Flash memory chip of the firmware, whereas any removable media device is generally designated as D: or E:. It is worth mentioning that every access from a client to file server (F32) takes place via a file server session, by means of RfS server session class, which implements all the basic services to interact with the file system.[3] We can obtain information about drives and volumes, act on directories, obtain notification about the state of files and directories, analyze file names, verify the integrity of the file system, and finally, manage drives and file systems.

3.3 Platform Security

Platform security [6],[10],[11] on Symbian OS v9 prevents applications from having unauthorized access to hardware, software and system or user data. The intention is to prevent malware, or even just badly written code, from compromising the operation of the phone, corrupting or stealing confidential user data, or adversely affecting the phone network. Every Symbian OS process is assigned a level of privilege through a set of capabilities, which are like tokens. A capability grants a process the trust that it will not abuse the services related to the associated privilege. The Symbian OS kernel holds a list of capabilities for every running process and checks it before allowing a process to access a protected service.

There are four different types of platform security capability, when digital signing is considered. The differences arise because of the sensitivity of the data or system resources the capabilities protect, and the requirements that are placed on the developer before they are given permission to use them. The capabilities of a process cannot be changed at runtime. The Symbian OS loader starts a process by reading the executable and checking the capabilities it has been assigned. Once the process starts running, it cannot change its capabilities, nor can the loader or any other process or DLL that loads into it affect the capability set of the process. A process can only load a DLL if that DLL is trusted with at least the same capabilities as that process.

The Symbian OS file system is partitioned to protect system files (critical to the operation of the phone), application data (to prevent other applications from stealing copyrighted content or accidentally corrupting data) and data files personal to the user (which should remain confidential). This partitioning is called data caging. It is not used on the entire file system; there are some public areas for which no capabilities are required.

Directory		Capabilities			
		None	AllFiles	TCB	AllFiles+TCB
\resource	Read	✓	✓	✓	✓
	Write	x	x	✓	✓
\sys	Read	x	✓	x	✓
	Write	x	x	✓	✓
\private\<ownSID>	Read	✓	✓	✓	✓
	Write	✓	✓	✓	✓
\private\<otherSID>	Read	x	✓	x	✓
	Write	x	✓	x	✓
\<anyOther>	Read	✓	✓	✓	✓
	Write	✓	✓	✓	✓

Fig 4 DataCaging Capabilities

However, some directories in the file system can only be accessed using certain capabilities. Each Symbian OS process has its own private folder, which can be created on internal memory or removable media. The folder name is based on the Secure Identifier (SID) of the process. A SID is required to identify each EXE on the phone and is used to create its private directory. With the previous Nokia phone generations, for instance the S40

series, the logical acquisition of the device content was possible, by means of Symbian OS APIs, which was able to copy the entire file system content on an external memory device. With S60 the access is restricted by data caging, the figure 4 shows table with folder security access (data caging) based on the application capabilities.

Interestingly, the file system restriction policy is fully contained in the file known as SWIPOLICY.INI [1], located in the folder z:\system\data\ . The original policy file related to a Nokia Symbian based smart phone is illustrated as follows.

```
AllowUnsigned = false
MandatePolicies = false
MandateCodeSigningExtension = false
Oid = 1.2.3.4.5.6
Oid = 2.3.4.5.6.7
DRMEnabled = true
DRMIntent = 3
OcspMandatory = false
OcspEnabled = true
```

```
AllowGrantUserCapabilities = true
AllowOrphanedOverwrite = true
UserCapabilities = NetworkServices LocalServices
ReadUserData WriteUserData UserEnvironment
AllowPackagePropagate = true
SISCompatibleIfNoTargetDevices = false
RunWaitTimeoutSeconds = 600
AllowRunOnInstallUninstall = false
DeletePreinstalledFilesOnUninstall = true
```

It is interesting to observe that the capability set defined in the previous file is limited, and it restricts the interaction with the file system of the mobile platform. According to the standard documentation, the various parameters can appear in any order. Moreover, UserCapabilities set might be changed, by adding the required capabilities such as those illustrated in the following modified version of policy file.

```
AllowUnsigned = true
MandatePolicies = false
MandateCodeSigningExtension = false
Oid = 1.2.3.4.5.6
Oid = 2.3.4.5.6.7
OcspMandatory = false
OcspEnabled = true
AllowGrantUserCapabilities = true
UserCapabilities = AllFiles DiskAdmin NetworkSer-
```

vices

```
LocalServices ReadUserData WriteUserData
UserEnvironment MultiMediaDD NetworkControl
CommDD ReadDeviceData WriteDeviceData
SISCompatibleIfNoTargetDevices = false
AllowRunOnInstallUninstall = true
AllowPackagePropagate = true
DeletePreinstalledFilesOnUninstall = true
```

The illustrated policy file can be written directly in the original firmware of the phone and, subsequently, up-

loaded by means of re-flashing. As a result, the complete C: disk content might be collected, with standard self-signed APIs, and thus analyzed, to extract the full set of probatory data which might be usually found on a mobile platform. This is the usual approach for other mobile platforms as well, where the primary image, the one which contains the entire set of evidence, can be obtained without any restrictions. Unfortunately, such a scenario is far from the reality, and we need to evaluate others ways to access the digital data content of the smart phone.

So far, if an application needs to have the complete access to the phone file system, it has to be authorized by means of the Symbian signing procedure with AllFiles capabilities, which requires a special certificate released by TC Trust Center. Three steps are required to sign an application. Initially, the installation file generator, make-sis.exe, creates the installation files (extension.sis) from information specified in the package file (extension.pkg). After that, if the application is for the international market, it will be signed with an ACS Publisher ID, by means of the Symbian Signed service. Conversely, it will be signed with a user-generated certificate, which might be created with makekeys.exe. Finally, the Installation File Signer (signsis.exe), digitally signs the installation files with the proper certificate, by generating, as a result, a .sisx file.

4 PROPOSED DATA COLLECTION SCHEME FOR SYMBIAN SMARTPHONE

This paper suggests a distinct approach both in development and extracting of data from the device. Application is developed as a composable part —A part provides services to other parts and consumes services provided by other parts. Each plug-in exposes a contract identifier so that it can talk to other parts in the application. The data extracting is based on client server architecture. During acquisition, the tool should have full access to the object store, as discussed there are very severe constraints to obtain a full physical image, so a logical copy of the object store is proposed. The client part is installed on the desktop PC the server part is copied into the mobile device giving us API access to extract a copy. The first problem that had to be tackled is the deployment of the tool onto the device. A number of ways to place the agent-based tool on the device were considered. The tool can be packaged as a SIS installer, so it can be sent to the phone using Bluetooth, infra-red or file transfer using the PC suite. A SIS file is a special software installer for the Symbian platform. Even though it may change certain parts of the file system, the changes are very little.

4.1 Data Acquisition

Data acquisition is the major step in forensics process. According to the forensics principle, the original data cannot be

used for any forensics analysis. So we need to create a copy of the logical data present in the mobile device. This is achieved by developing an agent, which is to be installed on the target device. The module uses Symbian SDK, AT Command Set (SIM card commands), FBUS and Connectivity SDK to read the file system. The module is capable of establishing a connection and exchanging data with an externally connected host computer. Figure 5 shows setting screen of the acquisition GUI which allows the investigator to select the phone model, which further enables or disables the features available.

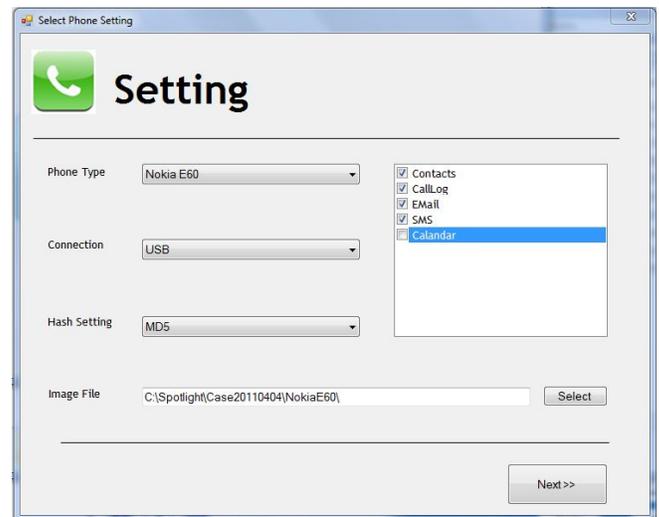


Fig 5 Symbian Acquisition Process

Since the tool uses standard APIs to access the file system and uses an agent, which is the code signed by the signing authority, we can reasonably believe that these APIs will not change the device content. The main issue with this approach is that it requires a piece of software, called the agent, to be loaded on the device to acquire the content.

4.2 Data Analysis

The tool creates a logical copy of the data present in the mobile device as a file at the desktop PC, where the client programme is running. The tool also supports to generate the hash value, which will prove the authenticity of the acquired data. The image created can be loaded in an analyzer so that the data present in the mobile device can be viewed for further analysis. The tool provides important forensic information like Contacts, Call logs, SMS etc. This information will help the investigation agencies to get some cues for further investigation. The analysis tool shows all these information in separate file viewers. The incoming outgoing and missed call details are displayed separately. Also the Inbox, Outbox, Draft, Sent and Deleted SMS are categorized in separate viewers. The analysis tool is also provided with keyword and file search facility, which is the key feature of a forensics tool. User can add any keywords and file extensions in the box provided and the tool will search the entire

image for the string entered. It shows the search hits in a separate viewer.

5 CONCLUSION

To summarize, standardizing the process of digital forensic methodologies for mobile phones are still in their infancy stage, the kind of data we need to look for, the security paradigm are new. As the platforms evolve and mature we should see more robust imaging tools e.g. VMWare tools for Android platform. For now, the hardware approach seems the only one which should be able to give a bit-by-bit image of the flash memory content thus preserving the content of the investigated phone. But software approach works for acquiring specific items. For instance; it is certainly possible to extract the entire set of probatory data, such as SMS, MMS, pictures, video clip and phone book, by using application APIs.

ACKNOWLEDGMENT

We would like to thank Mr. Thomas K L, Joint Director, at Resource Centre for Cyber Forensics (RCCF), Centre for Development of Advanced computing (CDAC) Trivandrum, for his valuable suggestions and support. This work was done at the RCCF, CDAC, Trivandrum, Kerala, India.

REFERENCES

- [1] Symbian-Ltd. Symbian OS library for application developers. Available at: <http://www.symbian.com>.
- [2] Morris B. The Symbian OS Architecture Sourcebook: Design and Evolution of a Mobile Phone OS. John Wiley & Sons, Ltd, 2007
- [3] Michael Aubert, with Alexey Gusev ... [et al.], Quick Recipes on Symbian OS, Mastering C++ Smartphone Development. John Wiley & Sons, Ltd, 2008. pp. 529-551.
- [4] Svein Yngvar Willassen, Forensics and the GSM mobile phone system, The International Journal of Digital Evidence, Spring 2003, Volume 2 Issue 1
- [5] Richard Harrison, Mark Shackman, Symbian OS C++ for Mobile Phones Volume 3. John Wiley and Sons, Ltd. pp. 204-206
- [6] Michael Aubert, Quick Recipes on Symbian OS Mastering C++ Smartphone Development, John Wiley and Sons, Ltd. pp. 60-63
- [7] Breeuwsma M., Jongh M. D., Klaver C., et al. Forensics data recovery from flash memory. In Small Scale Device Forensics Journal, 2007, 1.
- [8] Nielsen 2011, Who is Winning the U.S. Smartphone Battle?[Online]Available: http://blog.nielsen.com/nielsenwire/online_mobile/who-is-winning-the-u-s-smartphone-battle/
- [9] Dan Frommer 2011, businessinsider.com, RBC Capital Markets CHART OF THE DAY: 99.7% Of People Still Haven't Bought A Tablet Yet [Online] Available: <http://www.businessinsider.com/chart-of-the-day-heres-how-huge-the-tablet-market-could-get-2011>
- [10] Jo Stichbury Symbian OS Explained, , John Wiley and Sons, Ltd. 2004, 1.
- [11] Iain Campbell, Symbian OS Communications, John Wiley and Sons, Ltd. , 2007.
- [12] OMA(2001).Syncml sync protocol. Technical Report 1.0.1, Open Mobile Alliance
- [13] Symbian-Ltd. Carbide.c++: Introductory White Paper. Forum Nokia, Version 1.1; 2007

-
- Deepa Krishnan is currently Assistant Professor in PIIT, Mumbai University, India ,PH-91 977342 3043 ,E-mail:deepa@pointingarrow.com
 - Sathesh Kumar.S is currently Scientist at CDAC Trivandrum, India,PH-91-9995090885,E-mail:satheshks@cdactvm.in

Optimizing water fountain as community based therapy for better health continuance – Preliminary Study.

Nasser Aly, Jahangir Kamaldin, Yohei Kobayashi, Eva Banninger Huber.

Abstract- Background: The healing aspects of natural sounds therapy including water sound are beginning to get a lot of interest in the medical fields. Numerous studies conducted on the value of natural sounds as a therapy for various psychological treatments. Objective: to observe the influence of water fountain melodies on anxiety level among the experimental groups before undergoing dental checkup and to compare the effect of different water fountain melodies in reducing anxiety level by measuring the changes of brainwave signals. Methods: 4 types of water fountain melodies have been selected from 100 created melodies for the purpose of lab test. In this preliminary study we analyzed Electroencephalograph (EEG) data of 12 subjects. Volunteers divided into 4 experimental groups each group listen to one type of melody and one control group with no melody. Results: statistical analysis showed that Melody 2 has the first ranking which means that Melody 2 give the most effectiveness to brain in reducing anxiety. Conclusion: Melody intervention may has noticeable effectiveness on the brain in reducing anxiety than no melody and different melodies may have different effectiveness.

Keywords: Natural sounds, Water fountain, Melodies, Electroencephalograph, Teenagers,

1Background

Music has a countless of health-related benefits, both psychologically and physiologically. However, because the healthful effects of music have not been fully explored scientifically, many questions about the efficiency of music have not been answered yet. Scientific attention begin to recognize this concept for alternative treatments especially for pain relief management, stress reduction, improvement of sleep patterns, development of general physical and mental wellbeing (Mullooly et. al, 1988). With the advancing

-----□-----

technologies such as EEG (Electroencephalogram) and ECG (Electrocardiogram), physiological effects of sound on human mind and body can be investigated and quantified (Abdul Kadir et. al, 2009). Many authors reported that more researches are needed to study the effects of music on patients' stress and anxiety reduction (Bradt J and Dileo C 2009).

Nasser Aly; Advanced Medical and Dental Institute, University Sains Malaysia

Jahangir Kamaldin; Advanced Medical and Dental Institute, University Sains Malaysia

Eva Banninger Huber; Institute of Psychology, University of Innsbruck, Austria

Yohei Kobayashi; Center of Policy and International Studies, University Sains Malaysia

A clinical trial had done on the effects of music therapy on female breast cancer patients' anxiety following radical mastectomy and concluded that music therapy has positive effects on decreasing their anxiety (Xiao-Mei Li et al 2011). The healing aspects of natural sounds therapy including water sound are beginning to get a lot of interest in the medical fields. Numerous studies conducted on the value of natural

Hypothesis:

It is predicted that water fountain rhythms could affect teenager's emotions in managing their uneasiness before undergoing dental checkup.

2 Objectives:

We aim to observe the influence of water fountain melodies on anxiety level among the experimental groups before undergoing dental checkup.

3 Methods:

Nearly 100 types of water fountain melodies have been created. 4 types of melodies have been selected from the collection for the purpose of lab test. 4 types dynamic white noise have been created one for each type of melody. The sound has been modified based on lab test results. Hands on training for EEG recording using Enobio.

Inclusion criteria

- a. No previous anxious dental experience
- b. Not having serious disease that can influence in EEG examination (e.g., brain disease)

Exclusion criteria

- a. If in a serious debilitating oral disease (acute periodontal pain, pulpitis, abscess, or other acute infections, attachment loss or gingival recession, root hypersensitivity that might cause tooth hypersensitivity)
- b. Patients with severe anxiety and non-cooperative

Once volunteer arrived, divided into groups by sequence of:

Volunteer 1 will listen to Melody 1

Volunteer 2 will listen to Melody 2

Volunteer 5 will not listen to any Melody.

sounds as a therapy for various psychological treatments (Coensel BD et al 2011).

Does water fountain melody reduce anxiety induced by dental checkup? Are there any significance changes between experimental groups with melody intervention and control group?

We expected that there are differences in reducing anxiety among teenagers after melody intervention in term of brainwave signals.

To compare the effect of different water fountain melodies in reducing anxiety level by measuring the changes of brainwave signals.

Recruiting volunteers were selected. Brochure and consent forms have been prepared and distributed for the volunteers to participate in this study. In this preliminary study we analyzed EEG' data of 12 subjects. The subjects aged 12-16 years old, will be selected according to the inclusive and exclusive criteria:

- c. Have no dentin sensitivity to air stimulation.
- d. No hearing problem

- c. Not interested in music
- d. Having medical or psychology disorder that might affect pain thresholds
- e. Using pain or anxiety medication

Volunteer 3 will listen to Melody 3

Volunteer 4 will listen to Melody 4

Volunteer's behaviors were not controlled by the experimenter as we would like to study volunteer's behaviors in the actual state before undergo dental checkup.

Discussion session to prepare the subject to be more serious during the experiment' steps.

In order to determine the effectiveness of water fountain melody in reducing anxiety, Melody Intervention introduced

Steps of Composing Raw EEG Data

Since EEG signals consist of robust data, permutation method has been selected for the analysis. Thus, exact significance levels will be obtained rather than approximations. EEG data was analyzed and the sampling frequency was set. Next, the intelligent signal processing techniques were developed in *MATLAB* to calculate EEG raw data. The artifacts were filtered

right before dental checkup. EEG recordings were not implemented during dental checkup session to avoid confounding variables, e.g. facial muscle movement will affect clean data collection. We decided that EEG analysis focus on Melody Intervention section to examine the effectiveness of water fountain melodies in boosting volunteer's relaxation and reducing their anxiety level before undergoing dental checkup.

by applying threshold value into the signal analysis program. Afterward, the computed data was separated into 2 minutes data for each condition. Lastly, *SPSS* was utilized for further statistical analysis.

4 Results

Distribution Assumptions

Descriptive Statistics

	N	Minimum	Maximum	Mean	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Ch1Mel	12	.81163	7.12011	2.4948779	2.030	.637	5.061	1.232
Ch3Mel	12	.99691	10.14306	3.0099535	2.297	.637	5.669	1.232
Ch4Mel	12	.91076	2.93909	1.7237240	.832	.637	-.031	1.232
Ch1Sil1	12	1.02284	5.47180	2.1807161	1.787	.637	3.487	1.232
Ch3Sil1	12	.78484	3.14936	2.0593842	-.400	.637	1.203	1.232
Ch4Sil1	12	.82166	3.19443	1.5003539	1.686	.637	1.682	1.232
Ch1Sil2	12	.70590	9.93326	2.4637347	2.426	.637	6.353	1.232
Ch3Sil2	12	.66646	3.00778	1.3134512	2.321	.637	6.653	1.232
Ch4Sil2	12	.68590	4.84021	1.6052200	2.043	.637	3.986	1.232
Ch1Sil3	12	.81660	2.70721	1.3177538	2.162	.637	5.300	1.232
Ch3Sil3	12	.55638	3.42722	1.4613246	1.449	.637	1.530	1.232
Ch4Sil3	12	.61847	9.91000	2.4178642	2.411	.637	6.166	1.232
Ch1WN	12	.54934	2.65079	1.2728990	1.054	.637	1.532	1.232
Ch3WN	12	.51279	4.93533	1.6131000	1.736	.637	2.748	1.232
Ch4WN	12	.63837	2.78854	1.2966301	1.418	.637	2.850	1.232
Valid N (listwise)	12							

Table 1: shows the descriptive statistics about the EEG data. In the present study, researchers have collected 12 subjects. Each subject was analysed by conditions/channels example melody/channel 1. From here, some of the data follows the normal distribution and some are not follows the normal

distribution since the value of minimum and maximum is not beyond -1.0 and 1.0 and it is permutation method, so we assume that this data will freely follows any distribution.

Behavior of Conditions

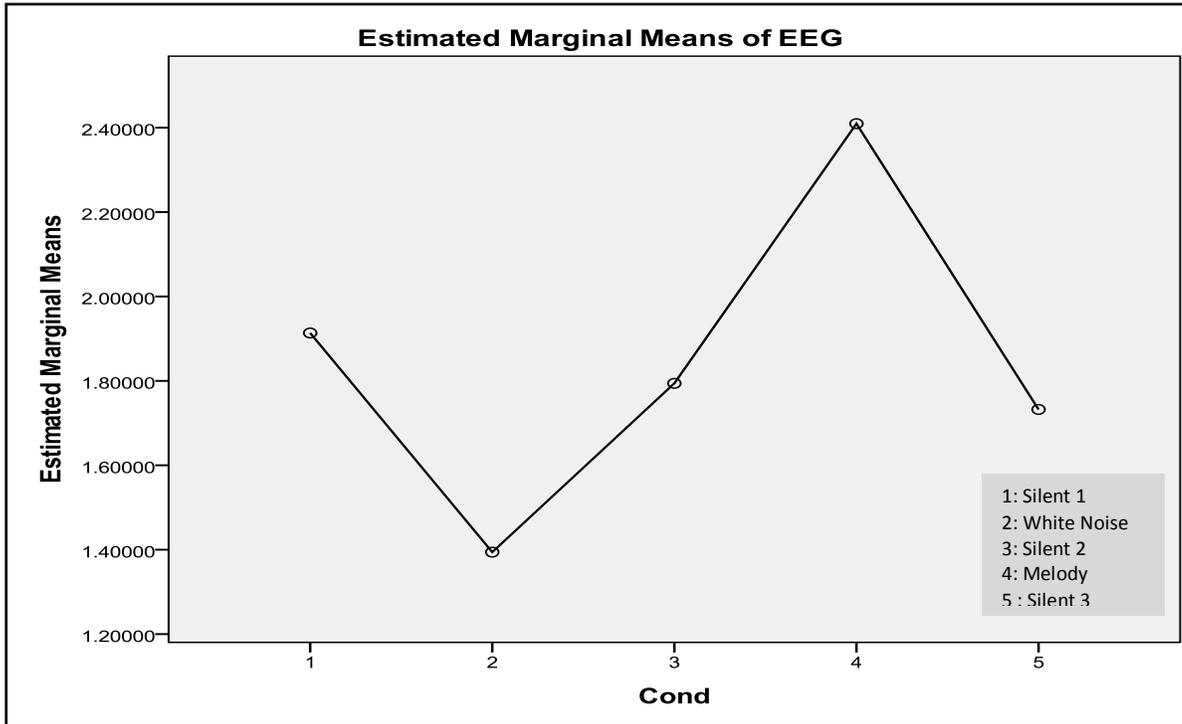


Figure 1: Estimated Marginal Means of EEG signal between Conditions.

Figure 1: shows the estimated marginal means of EEG data of all group of melody for each condition. Here, the purpose of this analysis is to see roughly the behaviour of EEG brainwave signals for each condition. From Figure 1.0, there is decrement of marginal mean from condition 1 (Silent1) to condition 2

(White Noise). Then, there are upward trend of marginal mean from condition 2 (White Noise) until condition 4(Melody). The trend starts to decrease from condition 4 (Melody) to condition 5 (Silent 5).

Behavior between Conditions & Channels

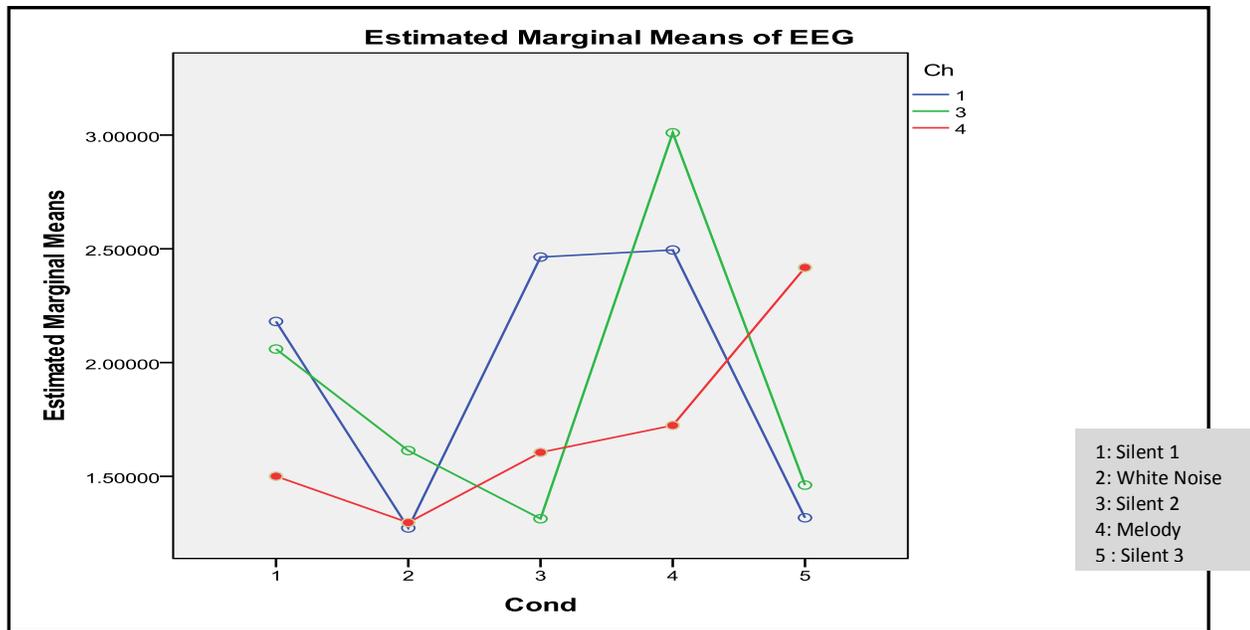


Figure 2: Estimated Marginal Means of EEG signal between Conditions and Channels.

Figure 2: shows the estimated marginal means of EEG signal between conditions and channels. The purpose of this analysis is to see which channel exhibits higher brainwave signal for each condition. From Figure 1.1, it shows that for condition 1(Silent1), the means of Channel 1 has the highest values followed by Channel 3, and Channel 4. For condition 2 which is White Noise, it seems that the mean of Channel 1 and 4 are same. While for condition 3 (Silent 2), the mean of Channel 1 has the highest values followed by Channel 4 and Channel 3.

For condition 4 (Melody), it seems that Channel 3 has the highest mean compare to Channel 1 and 4. Channel 4 for condition 5 (Silent3) has the highest mean, followed by Channel 3 and 1.

Choosing the Most Effective Melody Using ERP

Table 2 Means and Ranking of each Type of Melody Based on Channel

Type Melody		Ch1 Melody	Ranking Channel-1	Ch3 Melody	Ranking Channel-3	Ch4 Melody	Ranking Channel-4	Total Ranking	Ranking of Total Ranking
Melody 1	Mean	1.11329	5	5.56998	1	1.37666	4	10	3
Melody 2	Mean	2.31556	3	3.35199	2	2.31255	1	6	1
Melody 3	Mean	2.19476	4	1.93915	5	1.80915	2	11	4.5
Melody 4	Mean	3.62669	1	3.13635	3	1.80330	3	7	2
No Melody	Mean	2.80825	2	1.52450	4	1.23446	5	11	4.5

Table 2: shows the means and ranking of type of melody based on channels. The purpose of this analysis is to illustrate which melody is effective in reducing anxiety level among participant during dental check-up. The researchers rank the mean for each channel based on the highest mean value. In order to evaluate which melody gives more positive effect, the researcher sum up the ranking of each channels. From column Ranking of Total

Ranking, it shows that Melody 2 has the first ranking which means that Melody 2 is most effective among four melodies in reducing anxiety level. Followed by second ranking; Melody 4 and Melody 1 as the third melody that shows positive effects toward brainwaves signals. However, it seems that Melody 3 and No Melody have given the less effect on reducing the anxiety to subject before going to dental treatment.

5 Discussions:

- EEG is a study of charging electrical potential of the brain. Each electrode was placed on specific site of the scalp (T3, T4, T1) and measured brainwaves of different frequencies within the brain region (Abdul Kadir et. al, 2010). In our study, each melody has special characterization and accordingly different effect on brain activity (Fig. 3, 4, 5, 6). Changes of
- Figures 3, 4, 5 and 6 are the power spectrum of the part of each data Where Average dB is the loudness of each data. If it is 0, it means loudness is maximum and lower value means lower loudness relatively.

several rhythmic frequencies especially Alpha and Beta bands were examined to observe brain responses toward water fountain melodies. By analyzing brain signal activities, it is assumed that increasing alpha band power during listening water fountain melody will reflect relaxation in volunteer's conscious conditions.

Surround dB is the stereo width of the data. If it is 0, it means the data is narrow perspective and higher value means wider perspective

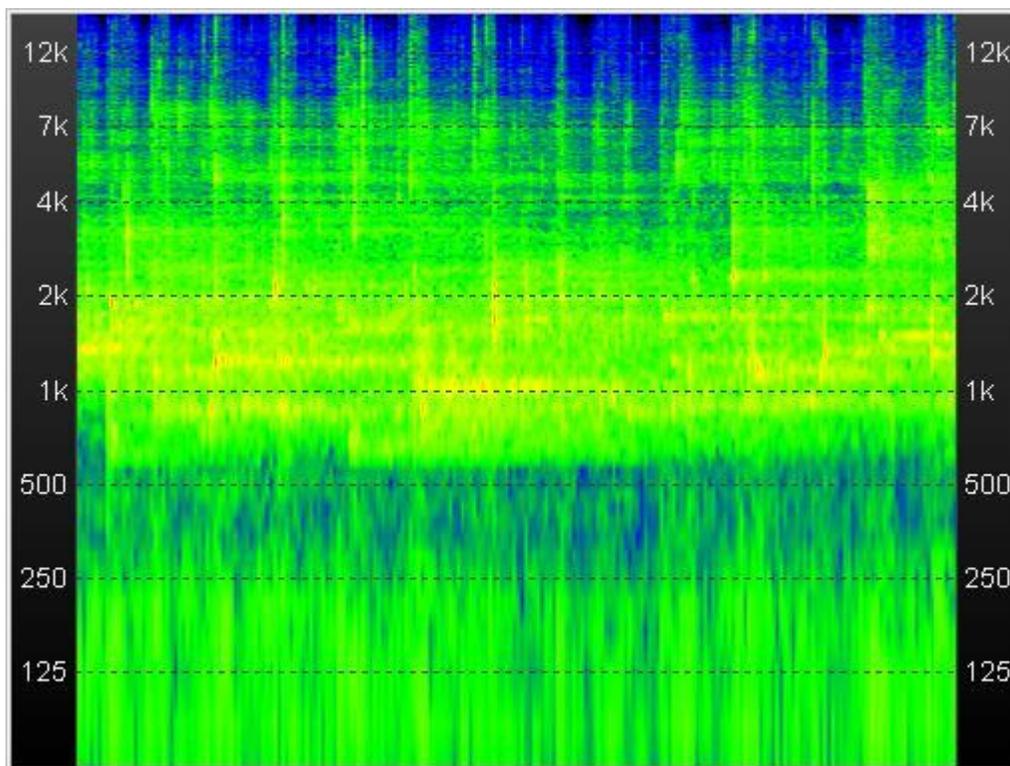


Figure 3: For Melody 1 Average dB -23.40 and Surround dB 1.97

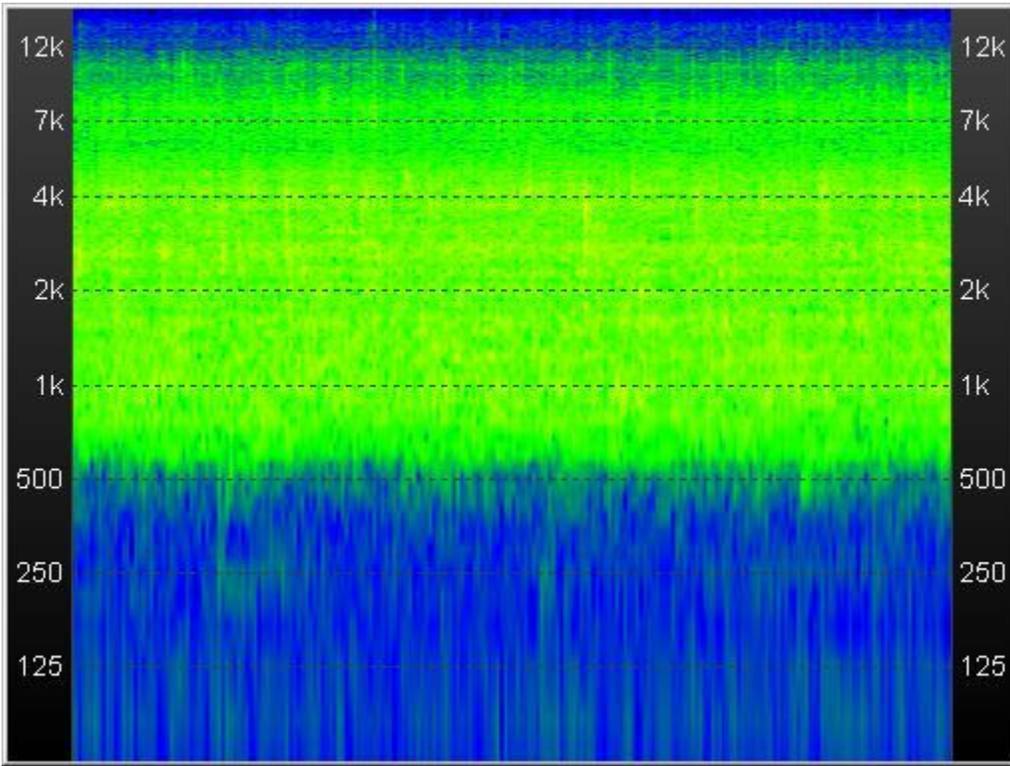


Figure 4: For Melody 2 Average dB -25.58 Surround dB 0.20, for Melody 3 Average dB -34.60

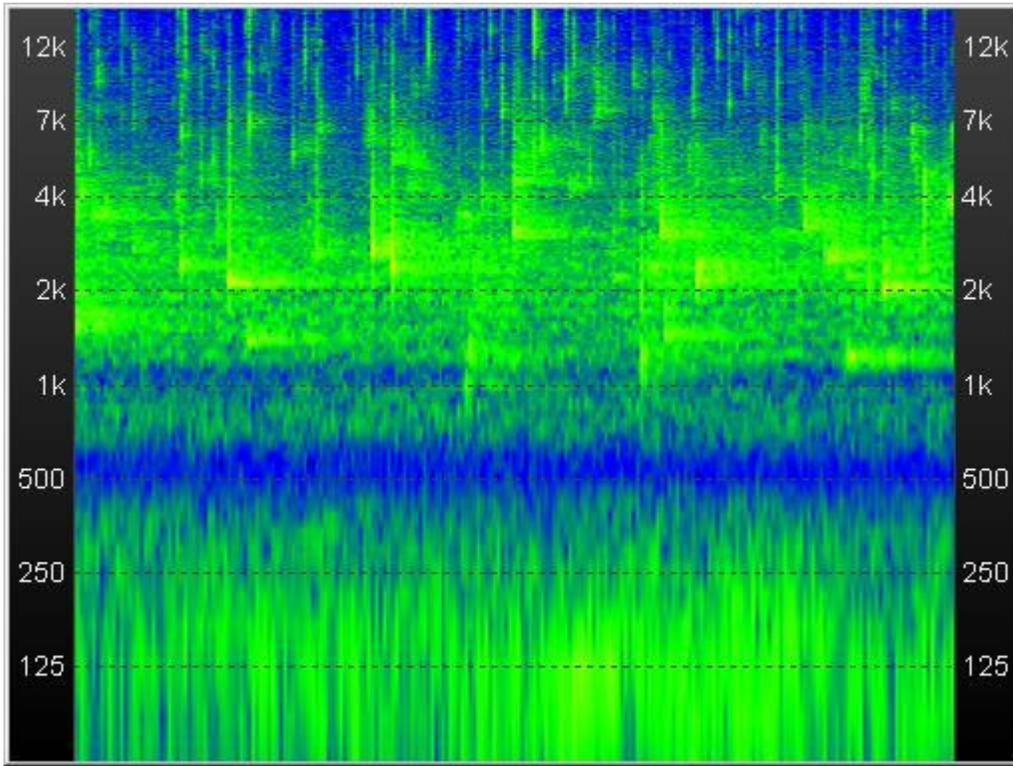
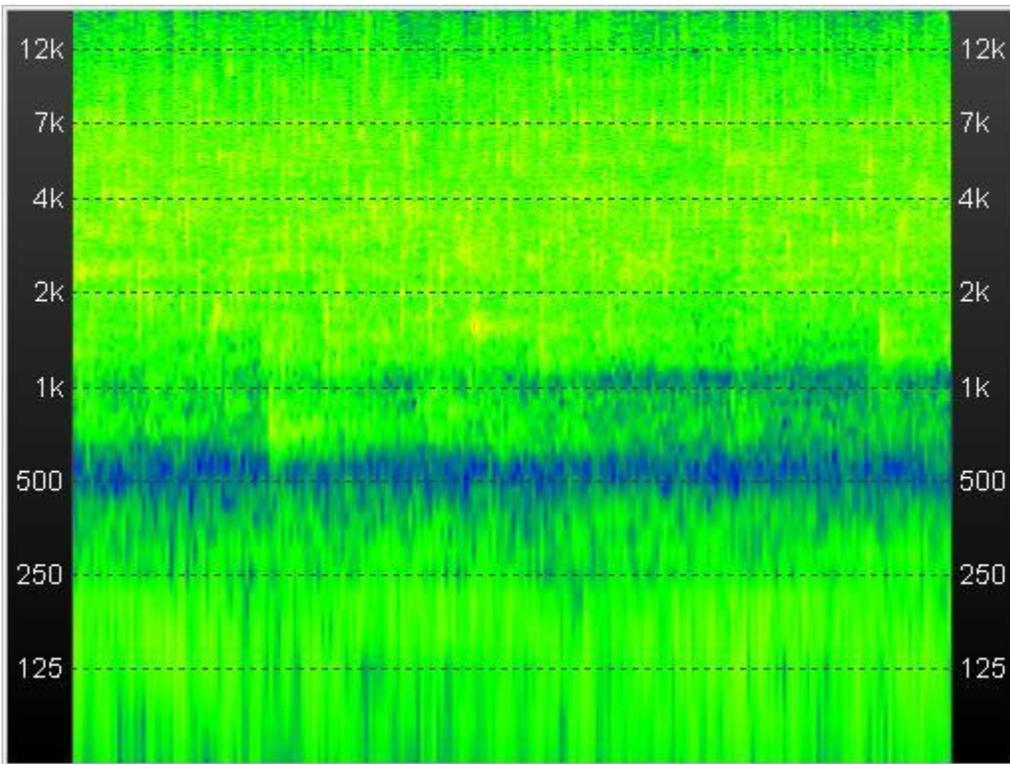


Figure 5: For Melody 3 Average dB -34.60 Surround dB 6.13



• Figure 6: For Melody 4 Average dB -21.82 Surround dB

5 Conclusions

Analysis of EEG collected data indicated different changes of the brain wave signals after melody intervention. They showed that melody intervention has noticeable effectiveness on the brain in reducing anxiety than no melody and different melodies have different effectiveness. Within the study limitations, our findings do not support that there are

significant differences between different melodies or between melody and no melody. Future analysis of data of large sample size should help to have more investigations of the changes in brainwave signals, facial actions and emotional regulations after melody intervention and provide more powerful statistical implementation.

6-Reference

- Abdul Kadir, R. S. S., Ghazali, M. H., Murat, Z., Taib, M. N., Abdul Rahman, H., & Mohd Aris, S. A. The preliminary study on the effect of nasyid song and rock music on brainwave signal using EEG. *Paper presented at 2nd International Congress of Engineering Education* 2010.
- Abdul Kadir, R. S. S., Yahaya Rashid, A. H., Abdul Rahman, H., Taib, M. N., Murat, Z., & Lias. S. Analysis of correlation between body mass index (BMI) and brain wave using EEG for alpha and beta frequency band. *Paper presented at 5th Colloquium on Signal Processing and Its Applications (CSPA)* 2009.
- Bradt J, Dileo C. Music for stress and anxiety reduction in coronary heart disease patients. *Cochrane Database of Systematic*. Published by John Wiley & Sons, Ltd 2009
- Coensel BD, Vanwetswinkel S, Botteldooren D. Effects of natural sounds on the perception of road traffic noise. *J Acoust Soc Am*. 2011; 129: 148-53.
- Mullooly, V. M., Levin, R. F., & Feldman, J. D. (1988). Music for postoperative pain and anxiety. *Journal of the New York States Nurses Association*, 4-7.
- Tilton, S. R. (2008). Review of state trait anxiety inventory (STAI). *NewsNotes*, 48 (2).
- Xiao-Mei Li, Kai-Na Zhou, Hong Yan, Duo-Lao Wang & Yin-Ping Zhang Accepted for publication 23 July 2011

Iran's Legislative Criminal Policy in Prevention from Crime

Fatemeh Nouri

Abstract— The society must be immune from anti social actions, while these actions are performed, i.e. general officials should do some actions that prevent individuals from committing this action. These strategies are called preventive actions. Criminal policy, which has involved control methods, was restricted to use oppressing method for a long time, but following this method's failure, tendency to preventive actions taking was generated. The goal of this research is to determine preventive criminal policy in Iran's laws and regulations. The duty of preventing from crime has been undertaken by judicature in the constitution but next to it some references undertake this duty in customary laws that this shows that it is essential to form a coordinated institution in order to prevent from crime.

Index Terms— Prevention from crime occurrence, Legislative criminal policy, Dangerous state, crime repetition, correction and rehabilitation, treatment.

1 INTRODUCTION

PREVENTION from crime and lawbreaking has been always one of the most important issues in human societies, which has been realized through imposing criminal, social and cultural policies.

In today's world, governments have not been very successful in prevention from crime and lawbreaking because lawbreaking and increasing crime have become a universal phenomenon and all of the world's countries have been engaged in it somehow, so, the necessity of fading out individuals' criminal motivation and decreasing lawbreaking severity in society resorting to preventive actions from crime is felt.

Framework determination of prevention from crime is very difficult from meaning and concept point of view. The word "prevention" has been defined in the meaning of removal, prevent, hamper and prior hamper. (Dehkoda glossary).

Some have defined crime preventive word as: Prevention from crime is performing precautionary actions in order to prevent from unwanted events (Fallah Babaei, 1: 2004-2005).

Criminologists consider prevention from crime as each contrivance and action that is applied in order to law-breaking inhibition and crime causing factors removal (Rajabi Pour, 19-21:2004-2005).

But this point is important that the first reaction that the human societies have performed as prevention from crime action have been oppressing actions, which was considering as only struggle method against crime and lawbreaking phenomenon for a long time. This oppressing action was the same as punishment, which in its own kind was imposed as hard and rough. Therefore, impos-

ing hard and rough punishments against crimes in order to prevent from crime commitment was among the specifications of old societies' criminal policy (Ardebili, 118:2007-2008).

But criminal policy in the field of combating the law breaking phenomenon was changed gradually with the appearance of sciences such as criminology and using their data in imposing criminals' punishment, especially criminals who have dangerous state, with this reasoning that punishment imposing not only effects on criminals' body and mind but also face their future life with various difficulties in society and punishment imposing must be used as the last alternative and weapon against lawbreaker and punishment substitutes must be used more (Noor Bahae, 528:1999-2000).

In this paper, we intend to review Iran's criminal policy about prevention from crime with this base that firstly which is the position of prevention from crime in Iran's criminal policy and secondly what are the modern methods of prevention from crime in Iran's criminal policy, but according to this fact that board of governors considered the only punishment imposing as prevention factor from crime occurrence in the begging and later Iran's criminal policy in this regard gradually was being moved towards using sciences data such as criminology, therefore, first combating crime methods in the past and then modern methods are reviewed and finally Iran's laws processing difficulties are explained about prevention from crime and necessary solutions are suggested in order to prevent from crime occurrence.

2 PREVENTION METHODS FROM CRIME IN THE PAST: THE MOST NOTABLE METHODS ARE AS FOLLOWS

2.1 Talion

Execution of talion order was giving to murdered heirs

• Department of Law ,Payame Noor University ,Po Box 19395-3697.ir.ofiran

in order to prevent from crime occurrence in the past. For instance, while a woman in Kashan has murdered another woman and her crime has been proved by "Amirza Mohammad" canonically, they ordered the murdered person's relatives to talion the murderer and murder her (Iran newspaper, 1870-71, No.: 232).

2.2 Plastering or putting between pillars

This is one of the important punishments of the Qajar dynasty period, in this manner that "they were generating two meters hole as much as the number of condemned persons next to the main roads or citadels and flags which were subjected to the general sight and were putting the condemned in it and were pouring some soft soils on their head, so that they covered above the condemned's breast and except the condemned's head and neck, his/her entire body was hiding in plaster then they were leaving the condemned and were going to him/her after two days and some of these condemned persons who had been alive after two days and were asking water, the executioner were cutting off their heads" (Wilz, 245-6:1989-90) and also refer to (Karlaseria, 126:1983-84).

2.3 Artillery or bombarding punishment

In this case, they were putting the condemned in the mouth of the cannon in front of the people's sights and after whipping them on their back they were firing the cannon and so the condemned was becoming disjointed. The first case of this punishment's form is related to Qajar Fathalih Shah Period, which was executed about Keliabee Khan who had rebelled against him. (Hedayat, 7469:2001-2).

2.4 Candles to decorate

It was among other methods of punishment that was executed especially about three groups of Naseradin Shah's opponents, in this case, they were puncturing the condemned's body and put candles in them and were taking the condemned around in alley and market in this position (Poolak, 242:1989-90).

2.5 Fastening with nail

This punishment was applied for dangerous condemned persons and veteran thieves in this manner that they were first pounding both legs and then both hands of the condemned to the wall by long and steely nails and then they were nailing three large nails from back to his/her body so that his/her breast is adhered to the wall and he/she was being kept in this position until his/her death (Wilz, 246:1989-90).

In addition to these punishments, other punishments were also imposing for dangerous criminals in the past such as shoeing, hacking, fining, hanging, taking eyes out, cutting ear and nose and exile or exilement..... (Karzan, 593:1994-95) and also refer to (Saeedi Sirjani, 223: 1997-98).

3 MODERN METHODS OF PREVENTION FROM CRIME

Punishment phenomenon and its goals were evaluated after passing this period in order to prevent from crime and as a result the necessity of punishment execution with that severity and intensity was doubted and the society accepted a more moderate reaction in turn for crime. This reaction had the title of security actions at first and then security and educative actions and its basis was to adopt strategies in order to prevent from crime (Noorbaha, 529:1999-2000). Implementation method of these actions is different in terms of different countries' criminal policy. In some countries, punishments and security and educative actions are executed together and in some countries, security actions are used after executing punishments (Saneae, 201:1974-75).

One kind of the above systems has been accepted in our country legislator criminal policy has been in order to fight against criminality inflation and prevent from crime occurrence (Noorbaha, 559:1999-2000).

Punishment is action or criminal action quitting reward in Iran's Law point of view and security actions are strategies that are used in order to prevent from crime because they are according to the dangerous position that the lawbreaker shows (Afrasiabi, 156:1998-99)

Security actions:

The first detailed law which has been approved about security actions in our country was security actions law approved in 2 may 1960 that includes three chapters and 21 articles (Formal newspaper, 1960-61: no.:4502).

Of course, there were some of the punishments such as exilement in general punishment law approved in 1927-28 which can be considered as a kind of security actions. The legislator explicitly has pointed to security actions in corrective law of general punishment law approved in 1973-74 and prescribes in article 15: security actions are as follows:

"Deprivation from all or some of social rights, compulsory residence in determined place, prohibition from residence in determined place, deprivation from holding a determined job, closing institution, deprivation from the right of guardianship or custody or executorship or stewardship"

Prisons affairs and security and educative actions statute was approved in Iran in 1982-83 (formal newspaper, 10924:1982-83) and a new statute under the title of "prisons statute and regulations" was approved in 1989-90 (formal newspaper, 1989-90: no.:12915).

And finally in 1993-94, "legal and executive regulations statute was approved that is currently executable together with the law of security and educative actions approved in 1960-61."

Article 1 of security actions law defines these actions and predisposes:

"Security actions include precautions that are adopted to prevent from crime repetition in the case of dangerous criminals."

Security actions are being separated totally from punish-

ments by this definition because: firstly, the only goal of the security actions is to prevent from crime occurrence. Secondly, these actions are only adopted in the case of dangerous criminals.

The article 1 of the above mentioned has specified dangerous criminals with this title that: "Dangerous criminals are ones whose records and moral and spiritual characteristics and commitment quality and committed crime make them subject to crime commitment in the future, whether they are legally responsible or not.

Dangerous state detection is the court's responsibility but no court can consider a person as criminal or among the dangerous criminals only by its own believe, rather the law provides some evidences for court to detect dangerous criminal (Ali Abadi, 131:1973-74).

Thirdly, according to the recent part of article 1 of the above law, sentencing security actions is allowable when a person commits a crime.

Fourthly, sentencing security actions is true about both responsible and irresponsible criminals. Responsible criminals are the individuals who are adult, wise, herald and free and on the other hand irresponsible criminals are individuals such as children, insanes, etc.

Some security actions about criminals according to law are as follows:

- Maintaining insane and mind disordered criminals in hospital
- Maintaining lawbreaker children in breeding and training center
- To surrender lawbreaker child to his/her parents and carers
- Maintaining addict criminals in addicts treatment centers
- Maintaining unemployed and stray criminals in agricultural and industrial workshops
- Closing institution
- Confiscation of dangerous things

The position of prevention from crime in Iran's legislative policy

Prevention from crime has legal legitimacy in Iran's law. Principle 156 of constitution has placed an appropriate action for preventing from crime occurrence in the domain of the judiciary duties. But with a look at normal rules approved by Islamic parliament and executive statutes of institutions and organizations, it is specified that prevention from crime has been mentioned among the duties of some organizations and institutions, which is because of the actions and duties that these institutions are responsible for in the field of fighting with criminality and social abnormalities (Zeinali, 112:2002-2003).

Among police force, the organs of prisons affairs' organization and also specialized polices, which were formed in recent years in police force such as anti-narcotics police, inform police, general security and information police, frontier guarding police and the most important of all is the prevention police, which propel the police force towards performing actions before crime occurrence by

taking the activity of police stations and outposts under authority, practically (Mirzaee Motlagh, "102:2002-2003).

4 Processing problems of prevention from crime in Iran

Prevention from crime has many problems; the most important of them are as follows:

- 1- Lack of a comprehensive in the field of prevention from crime
- 2- Organs' numerous, sporadic and causeless activities
- 3- Lack of cooperation between responsible agencies in prevention from crime affair and different positions of responsible organizations in prevention from crime
- 4- Poor partnership of people
- 5- Weakness of field studies
- 6- Restricted investment
- 7- Not using prevention specialists and experts on performing preventive actions
- 8- Budget inappropriate distribution
- 9- Ineffective system of criminal statistics and existence of inconsistent and invalid statistics
- 10- Lack of officials' enough attention to prevention from crime
- 11- Lack of preventive activity evaluation.

5 CONCLUSION

Legislative criminal law exhibits legislator Idea about crime subject. Tendency to compressor contrivances or preventive actions are being found in legislative criminal policy in the beginning.

In Iran's criminal policy, comprehensive and clear definition of prevention has been produced neither in laws nor in judicial procedure and our country has no coherent criminal policy in the field of prevention from crime.

According to the constitution's approval, judiciary as prevention affair's trustee should also accomplish an appropriate action resorting to all of the existent devices, facilities and powers. Generally, prevention from crime suffers from disorganization in Iran, for achieving such a big goal i.e. prevention from crime necessary programming must be performed and principles and techniques must be compiled, so that thanks to it Iran's legislative criminal policy in prevention from crime is realized.

It seems that the most essential action in order to prevent is paying attention to two important issues.

- 1- Formation of one institution, which is in charge of all prevention related affairs
- 2- Planning one comprehensive program for preventing from crime

The following techniques are also suggested to compile prevention comprehensive program:

- 1- Reviewing of effective causes and factors in generating crimes and their classification
- 2- Laws revision specially in penal section and compiling necessary laws according to the criminal policy principles

3- Revision of judicial system structure
4- To care and support family as main unit and responsible for primary social education of children and teenagers
5- To support education system and generating prevention of crime related specialized programs
6- Increasing international cooperation and supporting all of governmental and non-governmental organizations, specially about some crimes such as anti-smuggling
7- Using mass media in order to increase people's knowledge
8- Finally, evaluating the effect and efficacy of the related actions must be performed by evaluating and measuring that are the necessity of each program and according to that, revision of some programs is being performed.
It is necessary to mention that currently, judiciary judicial development supreme council plans to revise the laws and judiciary structure by quintuple commissions' formation in legal different sections. Penal laws and criminal policy compilation commission has taken numerous helpful and positive steps in this field by prevention committee formation, one of the actions of this committee is trying to adjust legal bill of prevention from lawbreaking parliament establishment. Meetings have been held about this and they plan to provide a comprehensive program in order to prevent from crime, it is hoped that good and helpful results can be acquired in the field of prevention from crime by applying all the available facilities and powers in the country.

REFERENCES

- [1] Afrasiabi, Mohammad Esmaeel: General punishment laws, volume 2, third edition, Tehran Firdausi publications, 1998-99
- [2] Ardebili, Mohammad Ali, General punishment laws, volume 2, fourteenth edition, Mizan publication, 2007-2008
- [3] Poolak, Yasakoob Edward: Iran and Iranians (travel account) translated by Keikavoos Jahandari, third edition, Kharazmi publications, Tehran 1989-90
- [4] Dehkhoda, Ali Akbar, Persian glossary, volume 13, Ofset Golshan publications, 1951-52
- [5] Rajabi pour, Mahmood, The procedure of social prevention from crime, Security knowledge quarterly periodical, fifth year, third number, 2003-2004
- [6] Zeinaee, Hamze, Prevention from lawbreaking, Social welfare quarterly periodical, sixth number, 2002-2003
- [7] Saeedi Sirjani, Ali Akbar, (by attempting), Vaghayea Etefaghie, reports collection of English secret writers in Iran's southern region, from 1870-72 to 1900-1901, third edition, Nikan publication, Tehran, 1997-98
- [8] Saneae, Parviz: General punishment laws, second volume, Tehran Iran's national university, 1974-75
- [9] Ali Abadi, Abdolhossein: Criminal laws, second volume, Mina publications, 1973-74
- [10] Security actions law approved in 1970-71
- [11] Fallah Babae, Hasan, Prevention better than treatment, papers collection of prevention from crime scientific-applied meeting, 2003-2004
- [12] Karlaser, Humans and rules in Iran (travel account) translated by Ali Asghar saeedi, first edition, Zavaar publications, Tehran, 1983-84
- [13] Kerzen, George, N: Iran and Its villages translated by Gholaam Ali Vahid Mazandarani, volume 2, four scientific and cultural publications edition, Tehran, 2004-2005
- [14] Mirzaye Motlagh, Ebadollah, Research thesis of police role in crime increase and decrease in Lorestan province, Naragh unit Islamic Azad University
- [15] Noorbaha, Reza: General punishment laws field, third edition, Dadafarin publication, Tehran 1999-2000
- [16] Wilz: Iran in one century ago (travel account), Translated by Gholamhossein Gharagoozloo, first edition, Eghbal publications, Tehran, 1989-90
- [17] Hedayat, Reza Gholi Khan: Rophtahosafay-e-Naseri history, by correction and annotation of Jamshid Kian Far, volume 15, first edition, Asatir publications, Tehran 1991-92
- [18] Iran newspaper: Thursday, 21 July 1870, No. 232
- [19] Formal newspaper: 1 August 1970, No. 4502
- [20] Formal newspaper: 5 July 1989, 12915
- [21] Wilz: Travel account, page 244 has an interesting description about one of the bombarding ceremony. And also Vaghayea Etefaghie: Thursday, 14 July 1864, No. 23, p 3

Iran Criminal Policy Due to Electronic Crimes

Fatemeh Nouri

Abstract— One of the most dangerous crimes that compromise socio-economical and political securities of the society is electronic crimes that have been manifested in the world era with its modern forms simultaneously with development in science and technology so the only condition for fighting against modern crimes and their prevention is to identify, analyze and recognize completely how they are committed. At the moment, in one hand there is no sufficient equipment in order to resist against crimes, on the other hand modern criminal policy does not have enough cognition and knowledge in order to fight against this new offenders. This study is aimed to explain concept of computer crime, identification of specifications and its different kinds and investigation of present legal solutions and indication of possible defects regarding high speed computerization of commercial-industrial and servicing activities in Iran.

Index Terms— Cyber crimes, informatics, communication technology, communicational networks, computer systems, criminal policy.

1 INTRODUCTION

BY entering to a meta- industrial society (informational), twenty's century has been faced with modern challenges in many life areas. This change has been occurred not only in the science but also in economy, policy and law so computer and communication technology have been followed by many professional discussions. Today, Information technology includes most of daily life especially in industrial countries; computer has been changed into an inclusive and essential tool in human lives and its usage in all fields is being increasing including scientific researches, medical affairs, finance, industry, and policy. Although, this technology has brought about countless advantages for governments and citizens, it causes many social, economical and judicial problems. Increase in computer abuses and imposed losses have made different people such as criminologists, lawyers and computer experts to study all aspects of this event. Of the most important and current computer crimes are as follows: fraud in credit cards, fraud in telecommunication lines, using computer for personal affairs by the personnel, non- authorized access to computer files due to curiosity, illegal copying from protected software and so on. This kind of crimes is being committed in a professional way that a considerable percentage of them cannot be identified. Today, judicial systems in the world in one hand face with growth of computer crimes and heavy losses remained from them and on the other hand, due to problems of legal regime present in countries, there is no power to prevent them effectively. It is while that everyone can commit this kind of crime against every citizen and maybe there is no crime such as this one regarding criminal motivation, kind of crime and its effective intensity. So it seems that the first step is to recognize the concept and topic. The main question is whether there is any

difference between traditional crimes and computer ones regarding degree, volume, quantity and device or there is a fundamental and natural difference. On the other hand, can it be said that only the tool and space have been changed? If the former is true, it will be possible to fight against computer crimes by principles and basics of current criminal policy. Theft, for example is theft but only the device of committing crime has been changed or in swindling, only the fraudulent device for carrying property has been changed and the nature of theft and fraud and how to commit them may be different in computer world. But if the latter is true, the current criminal system with present principles and strategies is not able to realize its objectives in direction of fighting with modern crimes. In this study, it is tried to analyze and study criminal policy regarding these crimes. In order to reach this goal, patterns and models of crime act will be discussed in two different spaces.

A: how to act a crime in a traditional model

One of the most important features of crimes in a real world is that in general a kind of positional proximity is necessary between criminal and victim. Theft, pocket picking, adultery, murdering and many other crimes have this feature meaning the criminal has to be close to the victim in order to act the crime. The second feature is that a criminal has a conflict with the victim while acting the crime and after crime completion; the offender can commit other crimes. His or her planning and goals focus on the crimes that he or she wants to act. Forger plans for forging a document and makes it then he acts other crimes. The third is that crime act in a real world is a function of physical limitations meaning that each crime even the simplest one such as pocket picking and street crimes needs initial readiness, planning and implementation. The more complicated the crimes, the more limitations. In simple words, traditional crimes have many limitations even after acting them regarding sale concealment and property transportation and so on. The fourth feature is that crimes are occurred in certain places of a city so

• Department of Law ,Payame Noor University ,Po Box 19395-3697.ir.ofiran

legal entities are able to centralize its forces and sources in possible places of crime act and show proper response and reaction against them. These features show that cyber crimes are as a new kind of crimes and have their own characteristics that distinguish them from other crimes. Later we will discuss them.

B: how to act crime in cyber crimes

In cyber crimes there is no necessary physical adjacent between criminal and victim. It is an unlimited crime. The criminal and victim may live in different cities even different continents. The only thing that a criminal needs is a computer system that is connected to an internet. By this simple equipment, he can attack victim's computer and commit a fraud or get personal information that enable him to commit criminal actions in a wide range by forging others' identity. In cyber world, the crime will be committed automatically by technology. This allows criminal to commit thousands of crimes with a very high speed without any legal threats and attention of the victim. So it is not possible to escape or take necessary actions that are existent in traditional crimes. On the other hand, criminal results and software and tools will be destroyed immediately after committing the crime so it is impossible to resist against them and state criminal law and procedure law face sever challenges in order to fight with these crimes.

2 PROBLEMS OF FIGHTING AGAINST CYBER CRIMES

1- In traditional crimes, pyramid model is used meaning report of criminal event, taking command for enquiry and prosecution, inspection and issuing sentence and executing punishments while in cyber space but it is impossible due to networked crime, speed of its committing and anonymous criminals

2- Lack of a certain borderline while acting the crime

3- Removal of evidences of crime act

4- One of problems in prosecuting and arresting criminals is that they are anonymous and unknown in cyber space. Face off and identity changes by makeup, plastic surgery, document forging existed in traditional crimes are not comparable.

5- increased crimes in cyber space is another problem because most of famous national and international companies do not want to show their unsafe business activities by disclosing acted crimes in order to not destroy their credits.

6- Cheap cyber crime: the most important device is a computer and telephone line in order to connect to the internet

7- Lack of international coordinated rules regarding definition of unit of cyber crime and judicial cooperation is one of the basic problems that needs long term international steps

By investigating these problems, it should be said that criminal policy has to have two main strategies to face cyber crimes.

3 STRATEGIES OF CRIMINAL POLICY

3.1 Reactive strategy

Reactive strategy means action taken after crime act which is inevitable and undeniable one in all crimes. It has no certain function in cyber world due to its unknown range, rate of crime act and problems existed on the way of discover, proof and judgment.

3.2 Strategy of prevention, control and management

Traditional crimes are a network organized for making human force and source but internet is a new social network that needs a modern criminal policy with a new preventive and controlling strategy meaning it requires active participation of all effective sections such as government, private sector and all people who are affected by cyber space. The most important preventive, controlling and manageable strategies for cyber crimes are:

1- Civil responsibility

All producers, importers, distributors and even users can be responsible for incorrect usage and production due to ignorance resulting in damage and losses to people of the society. Responsibility resulted from vicarious liability, citizenship responsibility and intensity of cause to preparation are not unknown entities in our legal system because negligence in making software tools and computer systems and not considering safety points can cause damages and losses to people, ethics and social credits (Katouzian, 374-1377:390) for example not considering safety points by a user and stealing of his device password can be an open window by which others' property can be stolen simply without identification of the criminal. So the user who has not been careful enough in protecting his device can be liable.

2- Criminal liability

The next step is to use vicarious liability in criminal dimension. The person who is able to control aggressive actions of others especially a person who benefits from those actions should be liable legally and criminally. It includes producers, makers, and designers of computer systems, providers of internet service and internet connection points.

4 Conclusion

We concluded in this study that today wide possibilities of computers not only have attracted good people but also criminals in a way that security forces face several crimes committed in internet network. There are many problems regarding prevention and persecution of these crimes and the current tools in criminal law cannot respond these problems so professional debates have been made in parallel with development of different kinds of computer crimes and international organizations have taken concentric actions in all countries in order to solve the problems but it is not enough so the most important thing for defending information networks against attacks

is to consider security points by increasing security of networks, installing protective systems and precise systems for discovering hackers and development of security software tools. In addition, all people working with computer especially judicial officers should be trained to cope with inspection, collection and protection of computer evidences. In general, it can be said that criminal policy effective for fighting against computer crimes and preventing them should be based on following principles and rules:

1- Supplying trained forces and establishing legal centers equipped with computer under control of the legislative power in order to discover crimes, collect evidences and enforce law of identification and prosecution of criminals.

2- Crimes of damaging behaviors in cyber environment

3- Crimes of initial actions

4- Informing companies and departments regarding ability of risk taking of computer systems and encourage them to use security measures

5- Promoting standardized security measures

6- Decreasing criminal situations and opportunities of using technical tools in committing an offence

7- Encouraging the victims to acclaim the criminal occurrence

8- Formulating proper rules and doing necessary amendment in current regulation of the country

9- resorting to international cooperation for discovering computer crimes and prosecuting criminals of these kinds of crimes. In order to resist against cyber crimes, the second liability is a better solution because sometimes those actions that do not cause damage naturally but can result in damage, should be recognized as crimes. For example, non authorized access to the system and information available in the computer should be placed in list of crimes because it is the first step in committing cyber offences against others and non- authorized access to information available in computer systems. This action has been recognized as crime in cyber offences convention in 2000 and English law of abusing computers in 1990.

REFERENCES

- [1] Ashoori, Mohammad, the book, introduction to internet and computer crimes, a new manifestation of criminal, Behnam press, second edition, 2007, p 13.
- [2] Katoozian, Naser, civil law in current legal order, Dadgostar press, 1998, Tehran
- [3] Ashworth Andrew, principles of criminal law clarendon press. Oxford 2000
- [4] Rosenberg Jonathan, cyber law THE LAW OF THE INTERNET, spriger USA 1997

- [5] Smith and Hogan, criminal law, ninth edition batter worth London, Edinburgh, Dublin 1999
- [6] Susan w. Brenner opcit.

Energy Efficient Lighting Control System Design For Corridor illumination

Jayashri A.Bangali, Arvind D.Shaligram

ABSTRACT:Energy efficient corridor illumination in commercial and residential buildings is becoming increasingly important in the energy conservation era. Lighting control systems with adequate algorithms are basically used for energy saving. The integration of daylight with electric light in building corridors during daytime hours is a key element in designing the lighting control system.

The most common form of electric lighting control is the on/off "toggle" switch. Other forms of lighting control include occupancy sensors, daylight sensors, clock switches, a variety of manual and automatic dimming devices, and centralized controls. The selection of sensor and type of lighting control depends on the application and the area where it is used. Mostly, occupancy sensor and photosensors are used in corridors. The corridor is not always occupied with people, rather people are moving continuously in a corridor. So, required illumination in a corridor is less as compared with a room or a hall. The twilight and night time illumination requirements are different and have additional behavioral and security issues associated. For corridors, required illumination or light intensity usually ranges from 50 to 100 lux [1].

The main aim of this paper includes the study and analysis of the good practice of energy-efficient lighting for corridor. The major findings in this study are that the energy-efficient lighting design could still be achieved without sacrificing the visual comfort and aesthetic requirement of the building, which is always a major issue of the owner. However, Energy-efficient lighting could be more easily implemented if it is considered right from the early design stage.

The proposed lighting control system is based on microcontroller and uses LED based luminaries as light source. The system includes patterning of luminaries for e in day and at night. This system uses photosensor and controls the illumination in a corridor depending upon the light intensity.

Indexed Terms: LED (Light Emitting Diode), Microcontroller, energy efficiency, Corridor illumination, patterning of luminaries, photosensors, energy saving, light intensity.

◆

1. Introduction:

The use of energy in buildings has increased in recent years because of the growing demand in energy used for heating, ventilating and air conditioning (HVAC) and lighting in buildings. Owing to the consistently growing demand, much effort has now been put towards reducing the demand for energy through energy efficiency in design. Energy efficiency is energy intensity, which, in simple terms, refers to the use of less energy to provide the same level of energy service or to do more work with the same unit of energy (EIA1 IEA2 Fickett et al, 1990). This objective can be achieved primarily by using a more efficient technology or process rather than by changes in individual behavior (Diesendorf, 2007).

Many studies have revealed that proper use of sustainable technology in lighting, such as the use of daylighting controls and low energy lighting, has a strong potential for reducing the demand for energy in commercial and industrial buildings (Busch et al, 1993; Nilsson and Aronsson, 1993; Min et al, 1997; Knight, 1999; Kim and Mistrick, 2001) and there is potential for improving the energy efficiency of lighting systems throughout the world (Mills and Piette, 1993). A recent survey of several companies has found that 23 per cent of all energy-saving opportunities could be achieved by improving the energy efficiency of lighting systems (BIE, 1996).

It has been reported that around 30–40 per cent of the total building electricity energy used in many commercial

buildings is consumed by the lighting systems (Swisher et al, 1994; Yarnell, 1995; Li et al, 2002). Research shows that a fair number of newly built buildings are still not designed in the way that the energy is used efficiently (Littlefair, 1996; Li and Lam, 2001). There are several instances in which lighting energy in the building has not been used efficiently. This could be because daylight is not efficiently integrated with the artificial lighting system, or in cases where integration does exist, energy savings using energy-efficient lighting technology have not been fully explored.

Enormous energy savings are possible using energy efficient equipment, effective controls, and proper selection of light source. Electric lighting design also strongly affects visual performance and visual comfort by aiming to maintain adequate and appropriate illumination while controlling reflection and glare. The Problem of energy saving and the achievement of visual comfort conditions in the interior environment of a building is multidimensional. Scientists from a variety of fields have been working on it for quite few decades, but it still remains an open problem. People spend about 80% of their lives inside buildings. So, achieving lighting comfort conditions in a building is very important and has direct implication to the energy efficiency of the building [2].The achievement of the lighting controller depends on its efficiency and properness in light level controlled illumination systems as well as types of light source used.

There are varieties of light sources such as Incandescent lamps, Reflector lamps, and Gas discharge lamps which includes Fluorescent tube lamps (FTL), Compact Fluorescent Lamps (CFL), Mercury Vapour Lamps, Sodium Vapour Lamps, Metal Halide Lamps

Tube lights/CFLs on average contain 4 mg of mercury, which has to be disposed by land filling. If a CFL is disposed off improperly the 4 mg contained within the lamp can get back into the normal ecosystem. Traditional 50-watt incandescent uses 438 kWh, which creates around 300KGs of carbon dioxide emissions per year. LED (Light Emitting Diodes) panel at 10watt consumes 5 times less power and hence will reduce CO₂ emission to 1/5th.

LED has its advantages such as a very low energy consumption level which directly contributes to savings on the light bill. LED has much longer life than traditional lighting (up to 50,000 hours). There is no infrared lighting, UV radiation from LED lights and also it doesn't contain mercury making it safer. Also LED has high luminous efficiency (at least 90-100lm/W), lower losses in the distribution of the controlled luminous flux compared to traditional lamps (emitting only a beam of 120°, whilst traditional have about 360°). LED is fully dimmable without color variation. However, LED lighting is not nearly as immediately affordable as traditional lighting. The cost problem becomes worse to change the existing lights as opposed to starting new.

2. Lighting Controls:

Lighting controls help conserve energy and make a lighting system more flexible. The most common light control is the on/off switch, Manual dimming, Photosensors, Occupancy sensors, Clock switches or timers and Centralized controls. Manual dimming controls allow occupants of a space to adjust the light output or illuminance. This can result in energy savings through reductions in input power, as well as reductions in peak power demand, and enhanced lighting flexibility. This type of technology is well suited for retrofit projects, where it is useful to minimize rewiring [3], [4]. Photosensors automatically adjust the light output of a lighting system based on detected illuminance.

Occupancy sensors turn lights on and off based on their detection of motion within a space. Some sensors can be also be used in conjunction with dimming controls to keep the lights from turning completely off when a space is unoccupied. These sensors can also be used to enhance the efficiency of centralized controls by switching off lights in unoccupied areas during normal working hours as well as afterhours [4], [5].

3. Control System:

Lighting controls can be grouped into two general categories: centralized controls and local controls. Centralized controls are used in buildings where it is desirable to control large areas of the building on the same schedule. For example in the morning the system can turn on lights a few minutes before the arrival of the employees. After the end of the working day the system can turn off lights again. During the day the lighting system can be adjusted in order to avoid peak demands (for example during noon at summer months). Localized controls are designed to affect only specific areas. There are various ways to control the lighting system for energy saving purpose. On-Off" switch is the simplest and the most widely used form of controlling a lighting installation. The initial investment for this set up is extremely low, but the resulting operational costs may be high. This does not provide the flexibility to control the lighting, where it is not required. Hence, a flexible lighting system has to be provided, which will offer switch-off or reduction in lighting level, when not needed. Grouping of light sources is also a way of lighting control systems for energy efficiency which can control light sources manually or by timer control. It can provide greater flexibility in lighting control.

Another modern method is usage of microprocessor/ infrared controlled dimming or switching circuits. The lighting control can be obtained by using logic units located in the ceiling, which can take pre-programme commands and activate specified lighting circuits. Advanced lighting control system uses movement detectors or lighting sensors, to feed signals to the controllers.

Whenever the orientation of a building permits, day lighting can be used in combination with electric lighting. This should not introduce glare or a severe imbalance of brightness in visual environment.

Additionally, sensor technology influences how well it detects occupants. Some sensors use infrared technology, some use ultrasonic technology, while others use a combination. There is a concern that some lamps have a shortened life if they are switched off and on frequently. Programmed start ballasts offer the softest and smoothest lamp start and are intended to improve lamp life if a frequent on and off occurs. [6]

The use of daylight in buildings is an important and useful strategy in replacing the need for high level of conventional energy for inside illumination. In this study, system is designed considering energy saving and lighting comfort together.

4. Study of the workspace:

The lighting control system for corridors uses daylight and electric lighting jointly to provide task, background or general luminance. The design of control system depends upon sky condition and solar location. In order to establish the lighting controls, luminance measurements are needed for a minimum of three different seasons representing winter, rainy and summer.

Here, illuminated indoor environment is the Corridor of Electronic Science Department, University of Pune. The area is first studied for deciding the optimum illumination level. For this study, the corridor in which the light control system has to be installed is divided into four sections A, B, C and D as shown in figure 1.1.

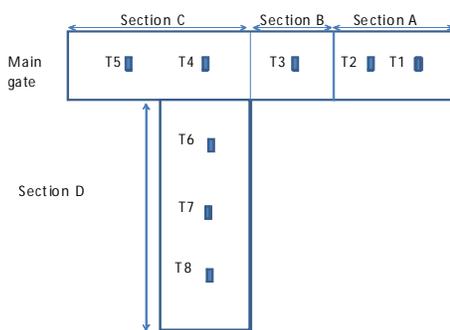


Fig1.1.Design of Corridor of Electronic Science Dept.

Existing lighting system contains 2 tubes in section A, 1 in section B, 2 in section C and 3 in section D. The corridor has Conventional Fluorescent Lamps/tubes/tube lights (T8) of 40 W. Fluorescent lamps are categorized according to their diameter. The requirements of illumination levels at different times in the corridor depend on the people working in that area. So a survey about the working hours of the people in that building has done which is given in table 1.1. To find out the minimum intensity levels in the corridor, several observations are taken at different timing of the day. Lux meter is used to take the observations.

Zone	Timing	No.of luminaires required
Fringe Hours	8 am – 10 am (in morning)	8
Working Hours	10 am – 6 pm	6
Fringe Hours	6 pm – 7 pm (in evening)	8
Non working Hours	7 pm – 8 am	4

Table 1.1

5. Control system design:

As the cost of energy has continued to rise, increasing effort has gone into minimizing the energy consumption of lighting installation. This effort has evolved along three major directions:

1. The development of new energy efficient lighting equipment
2. The utilization of improved lighting design practice
3. The improvement in lighting control systems.

While saving energy is of a great importance, there are some other associated benefits which should be considered. These are productivity and quality. However is quite difficult to quantify their influence. Lighting controls perform functions like on-off, time scheduling, dimming, and dimming due to presence of daylighting, lumen depreciation and demand control.

Lighting controls is an integral part of a lighting system. These controls must be responsive to the functional and aesthetic requirements placed upon it, and should perform these duties in an energy efficient manner. In general, there do not appear to be any general rules or guidelines that congenitally lead one to select specific controls. Energy savings due to daylight depends on climate conditions, building form and design and the activities within the building. In addition this factor is directly linked with the operating schedule of the building.

Providing daylight in a building does not by itself lead to energy efficiency. Even a well daylit building may have a high level of lighting energy use if the lighting controls are inappropriate. Case studies [7] have shown that in a conventionally daylit commercial building the choice of control can make 30-40% difference to the resulting lighting use. Wireless protocol such as ZigBee technology is also widely used in various areas for its excellent performance in reliability, power consumption, flexibility and cost for lighting control[8]. Lighting control could be performed by using fuzzy logic controller also [9]. Using motion sensors, lamp groups are turned on when motion is detected. Thus, energy consumption was prevented by turning off the lamp groups when there is no motion.

In order to assess the energy efficiency of lighting installation especially in corridor of a building, a criterion for the installed electrical power is proposed which is broadly applicable and easy to use.

6. Description of developed system:

The developed Lighting control system is basically based on energy efficiency of an indoor lighting installation for corridors which mainly includes use of daylight. The system is designed using 8052 microcontroller. The block

diagram of the developed control system is shown in figure 1.2.

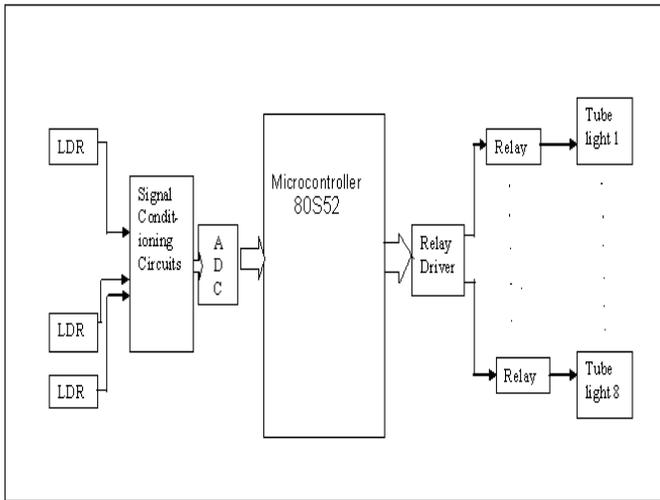


Fig 1.2 Block diagram of developed lighting control system

As shown in figure 1.2, the lighting control system uses microcontroller 80S52/80C52 microcontroller. The system has two modes. In mode 1, the inputs from the LDRs are considered. The resistance of the LDR is varying according to the light intensity in the corridor. The signal conditioning circuit converts it into voltage which is given as an input to the ADC (Analog to digital Converter). The output of the ADC is given to the microcontroller and depending upon the light intensity, the tube lights will be turn ON or OFF.

In mode 2, different options of patterning of tube lights are given to user. User can select any one of them and then accordingly tube lights will be On or OFF. User can also change or design the patterning of luminaires by making a particular tube light on or off just by setting it to 1 to turn ON or 0 to off the tube light. There are total five options are given in the system. However, they can be increased by changing the programming of microcontroller. The system uses relays which actually turn on or off the tube lights.

7. Results and Discussions:

After studying the illumination levels in all sections of the corridor (section A, B, C and D) and considering the comfort level of the occupants, it was found that minimum light intensity requirement in the corridor is about 60 lux. In section A, B, and D very less daylight reaches as these sections are interior part of the building. However, in section C sufficient day light is present because the main door of the department is open during the day time which

is in section C. So, accordingly, the patterning of the luminaires in the corridor is designed.

Energy saving due to LED tubes:

The eight existing light sources (T8 tube light of 40W) were replaced by 12W LED tube (which gives about 60 lux light intensity). Energy consumed by traditional T8 (40W) tube and by LED tube were compared. The calculations for both were done for 24 hours including day and night. Day is considered from 9am to 7 pm (10hours) and night is considered from 7pm to 9am (14hours) in the next morning. The formula used for calculation is as follows:

$$\text{Total Energy utilized (in kW-Hrs)} = \frac{\text{Total wattage of Luminaries}}{1000} \times \text{number of hours} \dots(1)$$

The calculations of energy consumed by T8(40W) tube light and LED tube light are done using formula (1) and is given in table 1.2.

	Energy consumed by 8 T8(40W) tube light	Energy consumed by 8 LED (12W) tube lights	Total energy saved by LED tube lights
Per day or for 24 hours	7.68 kWhr	2.304 kWhr	5.376 kWhr
Per Year or for 365 days	2803.2 kWhr	840 kWhr	1963.2 kWhr

Table 1.2

According to table 1.2, due to the use of LED tube lights 70% energy is saved compared to traditional T8(40W) tube light.

Energy saving due to Lighting Control System:

Further, the calculations of energy consumption with developed lighting control system are done. Considering the day light, patterning of luminaires are designed. The calculations of energy consumed in summer season, winter season and rainy season is performed and then total energy consumed for one year is calculated. According to season the day light varies. In summer season sufficient day light is present throughout the day till 6pm. However, in winter season because of cloudy climate less day light is present as compared to summer. In rainy season, assuming 56 days similar to winter season and 56 days similar to summer season and 20% days (that is 9 days) will be cloudy days. So the calculations for rainy season are different from summer and winter season. The chart of all the seasons with number

of days and availability of daylight in a year is given in table 1.3.

Season	No. of days in a year	Availability of light		
		Daylight	Twilight	No light
Summer	122 days	10 am – 6pm	6 pm – 7 pm	7 pm – 7 am
Winter	122 days	10 am – 6pm	6 pm – 7 pm	7 pm – 7 am
Rainy	56 days (similar to summer)	10 am – 6pm	6 pm – 7 pm	7 pm – 7 am
Rainy1	56 days (similar to winter)	10 am – 6pm	6 pm – 7 pm	7 pm – 7 am
Rainy2	9 days (cloudy days)	No sufficient daylight 10am – 7 p m		7 pm – 7 am

Table 1.3

According to the availability of daylight/twilight required number of luminaires varies which is given in table 1.1. Further, calculations of energy consumption as per seasons are calculated using the equation (1) and it is given in table 1.4.

Season	Energy consumed In Working Hours (kW-Hrs)	Energy consumed In Fringe hours (kW-Hrs)	Energy consumed In Night Hours (kW-Hrs)
Summer	70.272	29.280	76.128
Winter	70.272	29.280	76.128
Rainy1	69.696	29.04	75.505
Rainy2	6.912	29.184	5.616
Total	217.152	116.784	105.249

Table 1.4

As per table 1.2, the total energy consumption for one year without any lighting control system is 2803.2 units and using developed lighting control system is 439.185 units. So, total energy saving is about 84% per year by using developed lighting control system.

Conclusions:

There are several instances in which lighting energy in the building has not been used efficiently. This could be because daylight is not efficiently integrated with the artificial lighting system, or in cases where integration does exist, energy savings using energy-efficient lighting technology have not been fully explored. The major issues of the owners for installing the lighting control system is the initial investment, visual comfort of the people and aesthetic requirement of the building.

The main advantage of the developed lighting control system is that it can fitted in existing wiring setup and thus saved the initial installation cost of a system. The developed system is simple and cost effective as it is based on basic microcontroller. The daylight is integrated with the artificial light system which saves energy and LED tube light gives the aesthetic look to the building.

LED tube light which is known as the "green light source", saves energy up to 70%. In addition to it, LED lighting control system based on the microcontroller is the "Secondary energy saving" on the basis of LED light source which again saves energy and suggests many patterning options to the user.

Considering the comfort of the people and reducing the installation cost of the system, the developed system gives the simple and good design of a lighting control system for energy saving.

Reference List:

[1]. BS EN 12464-1:2002 Light and lighting - Lighting of work places - Part 1: Indoor work places CIBSE (Chartered Institute of Building Services Engineers) produces a Code for Interior Lighting

[2]. Alexandridis K., Dounis A. I., Supervisor of Digital PI-like Fuzzy Logic Controllers for Indoor Lighting Control in Buildings, eRA2, The Conference for the contribution of information Technology to Science, Economy, Society and Education, B.3.6. Greece, 2007, pp. 22-23.

[3]. Ryckaert W.R., Lootens C., Geldof J., Hanselaer P., Criteria for Energy Efficient Lighting in Buildings, Energy and Buildings, Vol. 42, 2010, pp.341-347.

[4]. Roisin B., Bodart M., Deneyer A., D’Herdt P., Lighting Energy Savings in Offices Using Different Control Systems and Their Real Consumption, Energy and Buildings, Vol. 40, 2008, pp.514-523.

[5]. Lighting and Daylighting, <http://www1.eere.energy.gov/buildings/commercial/lighting.html>

[6]. Slater A. "Lighting controls: an essential element of energy efficiency", BRE Information Paper IP5/87, Garston, UK.

[7]. Gary Steffy – Architectural Lighting Design, Second Edition – John Wiley & Sons, Inc. - New York (2002).

[8]. Maoheng Sun, Qian Liu, Min Jiang for An Implementation of Remote Lighting Control System Based on Zigbee Technology and SocSolution, ©2008 IEEE , ICALIP2008 978-1-4244-1724

[9]. SAFAK SAĞLAM, BULENT ORAL for Energy Efficient Lighting Controls and a Sample Application, ADVANCES in ENERGY PLANNING, ENVIRONMENTAL EDUCATION and RENEWABLE ENERGY SOURCES, ISSN: 1790-5095 ISBN: 978-960-474-187-8

Mrs. Jayashri A. Bangali,
Kaveri College of Science and Commerce, Pune, India,
e-mail: jayashri789@yahoo.com



J.A. Bangali born in 1974, is Asst. Prof. in the Department of Electronics, Kaveri College of Science and Commerce, Erandwane, Kothrud, Pune, India. She received her M. Sc. in 1996, M.Phil. in 2009 from University of Pune and pursuing her Ph.D. at University of Pune. She has more than 10 publications in the National and International Journals/Conferences. Her research interest includes smart space automation, lighting research, PC/Micro controller based Instrumentation and sensors.

Dr. Arvind D. Shaligram,
Department of Electronic Science, University of Pune, India



A. D. Shaligram born in 1960, is professor and Head of the Department of Electronic Science, University of Pune, Pune, India. He received his M. Sc. in 1981 and Ph. D. in 1986 from the same University. He has more than 70 publications in the National and International Journals. His research interest includes fiber optic and optical waveguide sensors, PC/Micro controller based Instrumentation and Biomedical Instrumentation and sensors.

Efficacy comparison of teaching problem solving skill and communication skills on pleasure and self-esteem in third boy students of Tonekabon city

Shahnam Abolghasemi, Hamidreza Vatankhah, Nikta bazleh, Samira Safdari, Hadi Moradi

Abstract— Aim of present study is Efficacy comparison of teaching problem solving skill and communication skills on pleasure and self-esteem in third boy students of Tonekabon city. Statistical society consists of 150 third boys students of Tonekabon city that have studied in educational year of 89-90. Among them 45 people had selected classified randomly in three 15 people groups (two experimental and one control group) as a sample. In order to data collecting, it is using oxford joyful (1990), Cooper Smith self-esteem questionnaire (1967) and these data were analyzed by using statistical methods of multiple co-variance analysis, Ben Freon alpha and pursuit test, that finally below results are obtained: teaching communication skills is more effective in students' self-esteem and joyfulness. teaching problem solving skills is effective in increasing students' self-esteem and joyfulness. Teaching communication skills are more effective in increasing joyfulness degree than teaching problem solving skill. there is no meaningful difference between efficacy of teaching communication skills method and problem solving skill on increasing students' self-esteem.

Index Terms— problem solving skill, communication skill, pleasure, self-esteem.

1 INTRODUCTION

ADOLESCENCE is one of the most sensitive and hardness life period that growing child should passes it and obtains necessary skill. This period, is converting period of childhood into adulthood. Psychologists called this period "third birth" due to its important and sensitivity (Abolghasemi, 1386).

Human beings required to functions for consistently resistance to stresses, different situations and life tensions that optimized them to understanding this ability. These functions are more fundamental and are shaped in the change process. The functions such as cognitions, emotions and behaviors that revealed in the structures forms like self-esteem, self- application, problem solving ability, resistance, self- understanding, social skills and understanding document(Noori Ghasem Abadi 1377, quoted by Yari, 1386). Skill means ability to do a work. Human beings are required to a series of skills for being success in every field. Such as skill in producing better product, skill

in education, gaining success and obtaining high scores. In training and education organization, skill means being familiar or access to knowledge. It is obvious that school cannot place its programs by emphasis on knowledge; it is not advisable to do this. Because knowledge mass has increased in a way that a long life is not sufficient for acquisitioning one field of sciences and techniques. Thus, training and education experts thought of teach students access way to knowledge rather than teaching knowledge and among this, considering capabilities such as thinking good, questioning truly, changing, observing and shortly find skill in problem solving were taken into account. Since by helping these skills we can solve many life issues. But in this way, there also was other point and that was learning living way. Because all thing is not shortened to lesson and knowledge, so school in addition to converting student into life learner, it also should learn him better living way and cooperation to others (Dadar, 1388).

From World health organization (WHO) point of view life skills are those skills which are taught to people for increasing psychological-social abilities and they enabled individual to face effectively with requirements and life challenges. Goal of teaching life skills is increasing psychological abilities and finally preventing from creating harmful behaviors to health and rising mental health level (Rahmat Zadeh, 1386).

Communication skills mean the ability to effectively communicate to others that both are useful for that individual and meet another person benefits. Experts believe

- Department of Psychology, Tonekabon Branch, Islamic Azad University, Tonekabon, Iran
- Department of Psychology, Tonekabon Branch, Islamic Azad University, Tonekabon, Iran
- Department of Psychology, Tonekabon Branch, Islamic Azad University, Tonekabon, Iran
- Department of Psychology, Tonekabon Branch, Islamic Azad University, Tonekabon, Iran
- Department of Psychology, Tonekabon Branch, Islamic Azad University, Tonekabon, Iran

that effective communication skills are necessary for success in today knowledge-based society. (Assar Zadegan, 1388).

Problem solving skill is well-organized and sequential approach that help person solves effectively his life problems. It is a behavioral, cognitive and emotional process in which an individual or group try to discover a solution or an effective confronting method for a special issue. Problem solving is reflective of one positive approach toward living that is counted the opposite of helplessness. Problem solving is a process, it is not a conclusion (Noor Mosavi, 1388).

In the psychological area, pleasure is a positive excitement. Pleasure is deeper than a temporary good mood. It is a constant feeling in positive excitements and understanding life reality. Pleasure can be defined as a constant pleasing feeling and an approach. Pleasure and mental health involve positive feelings such as enjoyment, comfort, being feeling and fascination in the life(Amani, 1385). Self-esteem is valuing, compliance, approve and confirm that a person feels about himself (Biabangard, 1373, quoted by Rezaee, 1381). Self-esteem is an assessment, assessing his characteristic that leads to child feeling about himself and his features. It may be very low self-esteem and poor self-imagination prevent children from talents objective evaluation and their skills, because their negative view which they have from themselves converts to a glasses that they see all of the information from its behind. Low self-esteem is also a big component in impairment period of weak action(Creshner, 1990).

The most direct intervention method, in order to rising social-psychological abilities, is those interventions that are enforced individuals resistant sources and personal abilities especially their social and mental capabilities. In regards to children and adolescents, this kind of intervention can be presented by teaching life skills in the supportive education environment that is school.

-in an investigation that conducted by Berlon and Macgil(1989), they found that when there is discomfort feeling and lack of pleasure, people' self-esteem decreases. In other words, there is a interaction between self-esteem and pleasure (quoted by Poorsanaee, 1387).

-Roznberg and et al(1995) have compared the effects of whole and educational self-esteem of 2213 students in tenth class. They showed that whole self-esteem had stronger correlation to joy (0/5) and negative emotion (0/43). But educational self-esteem was obtained more relationship to educational scores.

-Hghighi Pazhohesh and et al (1384), focused on examining the efficacy of life skills teaching on self-esteem and mental health in boys students of first grade in third school in Ilam city, results showed that life skills teaching leaded to increasing mental health and self-esteem of boys students in experimental group in comparison with testimony group.

-Poshtiban (1386) studied the problem solving style relationship to students' educational progression, self-esteem

and their way of thinking, the results showed that there is meaningful correlation between six constructive problem solving and self-esteem styles.

-in the Hemati investigation (1385) with the subject of "examining the efficacy of learning social skills on students' self-esteem and their friendship", this result was obtained that learning social skills caused increased self-esteem and decreased friendship in students.

-Poorsanaee in an investigation under the name of " self-esteem, self-imagination and internal control source relationship to excitement intelligence and pleasure in boys students of first grade third school in Bandar Anzali", found that there is a relationship between self-imagination, self-esteem and internal control source to pleasure, and pleasure amount is differ in internal control source from external control source. Also pleasure amount in positive self-imagination is differing from negative self-imagination.

-in Hemayat Talab and et al study (1384) that focused on investigating pleasure amount in athletic and non-athletic boys students in Tehran universities, they found that athletic group have self-esteem, satisfaction, efficient, positive temper, mental health and generally they have pleasure in comparison with non-athletic group.

-Shain Berger (2010), in his study with the subject of "positive psychological, optimism and situation", he showed that successful people have high self-confidence and they are optimistic. They used appropriate communication skills techniques in encountering with their around people.

-Elison, (1983) in his study found losing control feeling and creating individual dissatisfaction is one of the specific aspects of low self-esteem (Rezaee, 1381).

-Hashemi study (1374) about examining the social support, control source and self-esteem relationship to resistance methods in third school students in Azerbaijan Sharghi province, suggested that social support is one of the most important sources for confronting with stress. People, who have internal control source, give importance to their values and they are able to control the environment. More high individual self-esteem, using encountering strategies is more and most appropriate. More individual control source is internal, the individual have high self-esteem and social support. Received social support is dependent to individual self-esteem.

-Abedi studies results (1383) suggest that optimistic people have more pleasure amount than pessimist and realistic people, and optimism is one of the increasing pleasure methods (Moradi and et al, 1384).

-Esmaeeli (1380), focused on examining the efficacy of teaching life skills understanding in comparison with child training methods on students' self-esteem in Ardbil province, findings showed that average self-esteem of students who passed skills course was more than students who did not passed the above course. This finding was consistent to studies results of American psychological association (1994, 1996).

General aim of this study was comparing efficacy of teaching problem solving and communication skill on self-esteem and pleasure amount in students. Also primary hypothesis of the investigation presented as follow:

-there is a difference between efficacy of teaching problem solving and teaching communication skills on self-esteem and pleasure amount in boys students.

Sub-hypothesizes of the study:

1. Teaching problem solving is effective on pleasure amount of boy students.
2. Teaching communication skills is effective on pleasure amount of boy students.
3. Teaching problem solving is effective on self-esteem amount of boy students.
4. Teaching communication skills is effective on self-esteem amount of boy students.
5. There is a relationship between efficacy of teaching problem solving and communication skills on pleasure amount of boys students.
6. There is a relationship between efficacy of teaching problem solving and communication skills on self-esteem amount of boys students.

2 RESEARCH METHOD

Present study is a kind of experimental investigations that was conducted in the form of intervention. Study plan is pre-test and post-test with control group. Investigation statistical society consists of 150 boy students in second grade of third school in Tonekabon city who were engaged in study in 89-90 education year, among them and by using classified randomly sampling method 45 people were assigned to two experimental group and one testimony group that every group involved 15 people. In order to collecting data, it was used Oxford pleasure and Cooper smith self-esteem questionnaires. After gathering data, it is analyzed by SPSS statistical software, multiple co-variance analysis test, Ben Freon Alpha and pursuit test.

3 RESEARCH TOOLS

A) Oxford pleasure questionnaire (1990):

This scale created by Argyle for providing general size of pleasure in 1990 which involves five factors: satisfaction, positive temper, healthy, efficiency and self-esteem. Oxford 29 items scale is scoring based on four degrees spectrum from 0-3 and testees final scores collected. Final score will be at least from 0 to 87.

Alipoor and Harris (1386), in an investigation that they conducted, estimated 0/91 amount for kronbach alpha of this test.

B) Cooper Smith self-esteem test (1967):

Cooper Smith (1967) provide this scale based on overview which he performed on Rogers and Deymond' scale, and finally a questionnaire compiled with 58 items that its 50 items focuses on 4 sub-scales of family self-esteem, re-

search self-esteem, general self-esteem, social self-esteem and its 8 items are lie detector. Testees scores range are 0-50.

In Iran, stability coefficient of this scale has reported about 0/79 by Zare (1386).

4 RESEARCH FINDINGS

A) Descriptive findings:

Statistical characteristics of dependent variable components to self-esteem and pleasure in two experimental and control groups are presented in following table.

TABLE 1

variab les	Problem solving group		Communication skills group		Control group	
	mean	Standard devia- tion	mean	Stand- ard devia- tion	mean	Stand- ard devia- tion
Plea- sure	59/8 0	6/83	65/7 3	4/57	50/8 0	9/51
Self- es- teem	39/6 0	3/54	38/6 7	4/06	36/4 0	4/40

Regarding to the figures of table, there is no difference between the means of both groups of control and experi- in the pleasure and self-esteem's components, and these differences are an advantage for experiments group, and more than the control group.

b) Deduction findings:

It is used from multivariable co-variance analysis to test the major research hypothesis. The results of this analysis are as shown in table 2.

TABLE 2

MULTIVARIABLE CO-VARIANCE ANALYSIS, MIXED VARIABLE OF PLEASURE AND SELF-ESTEEM

Source	Value	F	Meaningful level	Eta
Mixed varia- ble of pleasure and self es- teem	0.144	30.23 6	0.000	0.620
pleasure		62.8	0.000	0.758
Self esteem		28.84	0.000	0.591

According to (table 2), the value of 0.620 suggests the effect value of performed training in mixed variable. The value of effects above 0.14 or (14%), showing a large effect. The results of above table suggest that there is a meaningful difference between the experimental groups

which have been under the influence of problem solving and communicating skills training and control group which has received no training.

$F(4, 74) = 30.236, P = 0.000, \text{Eta} = 0.620$

Regarding to pursuit test performed between two groups, there is a meaningful difference in means of two groups, and this difference is an advantage for experimental groups. Also about two groups which have trained the problem solving and communication skills, meaningful difference in increasing of pleasure variable, is an advantage for communication skills. While in self-esteem variable there is no meaningful difference. It means that effectiveness of both experimental methods in increasing the self-esteem is the same.

Sub-hypothesis 1&2: teaching problem solving and communication skills are effective for pleasure enhancing.

TABLE 3
THE REGULATING MEAN OF PLEASURE AND SELF-ESTEEM VARIABLES

De- pen- dent vari- able	Communication group		Problem solving group		Control group	
	mea n	stan- dard error	mea n	stan- dard error	mea n	stan- dard error
plea- sure	65.14	0.964	61.37	0.956	49.82	0.986
Self- es- teem	39.46	0.553	40.51	0.549	43.70	0.566

As shown in table 3, we observe that there is meaningful difference in enhancing the pleasure variable between regulating means of three groups.

$F(2,45) = 6.28, P = 0.000, \text{Eta} = 0.758$

So we conclude that teaching problem solving and communication skills affect the pleasure enhancing. The following test showed that there is meaningful difference between two training groups, and this difference suggested that the effectiveness of communication skills teaching in enhancing the pleasure is more than that of the problem solving expertise.

Hypothesis 3&4: group training of problem solving and communication skills affect the pleasure enhancing:

Regarding to the results of table 2, we have:

$F(2,40) = 28.8, P = 0.000, \text{Eta} = 0.591$

Problem solving and communication skills, thus, have been affected in enhancing the self-esteem. The results of the following test suggested that there is no meaningful difference about influence on self-esteem variable between two training groups, and their effectiveness in enhancing self-esteem is the same.

Hypothesis 5: There is difference between the effective-

ness of problem solving and communication skills teaching on pleasure value of boy students. There is meaningful difference between two training groups, suggesting the effectiveness of communication skills teaching in pleasure enhancing, and that the difference value in this group is more than that in problem solving skill group.

Hypothesis 6: there is difference between the effectiveness of problem solving and communication skills on value of boy students' self-esteem. There is difference between training groups in self-esteem enhancing, but it is not meaningful; and the effectiveness of training in self-esteem enhancing is quite the same and so the above hypothesis refused.

5 Discussion and conclusion

present paper was performed by purpose of comparing the effectiveness of problem solving and communication skills on boy students' pleasure and self-esteem, and following results was provided.

One of the results of this study was the effectiveness of problem solving and communication skill teaching on boy students' pleasure and self-esteem, that this finding is consistent to the findings of Tlado (1974), Gonzalez (1990), Darden (1996), Halford (2004), Hashemi (1374), Aghajani (1381), Iranpour (1382), Poshtiban (1386), Rahmati (1387), Poorsanaei (1387), and Sanaei Zaker (1384).

In relation to this finding, it can say that, undoubtedly, the teaching problem solving and communication skills, purposed in enriching the students' behavioral treasure, help the increasing of their pleasure and self-esteem, and that no training can completely success alone, regarding to the complex transaction of intrinsic tendencies and responses learned by human being.

What obtained from this research was influence of problem solving and communication skills teaching on the value of boy students' pleasure and self-esteem, which this result is similar to the findings of Khepg & Farnam (2004), Ebadi (1383), Bahari (1388), Jokar & Rahimi (1386), Moradi et al. (1386), Hilz & Argail (2001), and Abedi (1383).

To clear this result, it can say that solvent expertise entail the use of students and antecedent learned principles into a new structure. Equipping to ability and problem solving skill against the problems for human being, has a permanent influence in his achieving to peace. Also communication skills help people can to express their views, ideas, requests, needs, and thrills and that to request for help when they need. Equipping to ability of communicating expertise cause to reinforce the warm relationship with others and provide social and mental health more and more and increase the pleasure.

Also it was confirmed in further hypotheses of research, that problem solving and communication skills

have influence on students' self-esteem. This result is consistent to the researches of Rozenberg (2009), Khepk et al. (2007), Vichrovski (2000), Adib (1374) (quoted from Rezaei, 1381), Esmaeili (1380), Fathi (1385), Mehdizade (1388), Negat et al. (1386).

Difference between the effectiveness of problem solving skill and communication skill on the volume of students' pleasure was another finding of this study, specified that this difference is testimony group which has received the problem solving teaching, and suggested the major effectiveness of this training that confirms the findings of Kalt et al. (2005), Karor et al (2005), Rozenber (2009), Shokberger (2010), Jokar & Rahimi (1386), Bahari (1388), and Yoosefi Jooibari (1389).

The final hypothesis of research surveys the difference between effectiveness of problem solving and communication skill teaching on the volume of students' self-esteem and the result was that there is difference between training groups in enhancing the self-esteem, but this difference is meaningful and the effectiveness of teaching of both skill has actually same influence on self-esteem and refused this hypothesis. This result is consistent to Abnikifard's (1382). To support this consequence, we can say that despite the difference before and after the teaching problem solving and communication skill, statistically there is no such difference in value of self-esteem, that it can depends on understanding of receivers, their sense of need and personal features. Also establishing of problem solving and communication skill in individuals depends on growth periods, environmental and family structure, and congenital features, and may needs to short time training.

Successful learning of problem solving and contrasting skills, impress learner's sense about himself and others, and moreover, achieving to proficiencies change others' thought about individual. So achieving to contrasting skills change both personality and environment and this mutual element accelerate the promotion of psyche health. Reinforcing the ability of problem solving skill cause to increase the self-esteem and flexibility against the changes and promote positive thoughts to improvement and relying on self.

Whatever social knowledge progresses, it rather specify that the secluded, alone, and lake of communicating man subject to many social harms. It is through communicating that a person can express his/her feeling to others. It is possible just through safe communicating than Negative emotion could be controlled. Training of communication skills make people to achieve their value and perform positive behaviors throughout the life; and correlate to others in a way that these relations will have desirable effect and this desirability will cause to increase the duration of relations. While a non-effective relation cause to

misapprehension, led to dissatisfaction, sense of loneliness, and conflict in family and society and disorder the individual's self-confidence during the time, and so create social and mental harms and decrease individual ability to contrast the problems.

REFERENCES

- [1] 1. Abolghasemi, Shahnam (2008). Growth psychology, by emphasizing on mental health, Azad University of Tonkabon publications, first edition.
- [2] 2. Esmaeili, Mohammad (2000). Revision of the effect of training of life skills against necessities of breeding on students' self-esteem in Ardebil. Payam-e-Moshaver monthbook, third year, No. 6, p. 13-125.
- [3] 3. Poshtiban, Reza (2008). The relation of resolvent mode with studying improvement of students' self-esteem and their idea. Postgraduate thesis Islamic Azad University, Science-Researchs of Tehran.
- [4] 4. Pour Sanaii, Ghazal Sadat (2009). Study of relation of self-image, self-esteem, and source of internal control with emotion intelligence and merriness in boys student of Bandar Anzali. Postgraduate thesis of Common psychology. Azad University of Tonkabon.
- [5] 5. Hemayat Talab, Rasool; Bazaran, Sedighe; Lahmi, Reyhane (2006). Study of athletic and nonathletic boy students of Tehran Universities (Governmental and Payam-e-Noor).
- [6] 6. Dadar, Nasrollah (2010). Life skills, which skills? Moallem; training, analytic, and alarming month book of education ministry, cycle 28, No.244, P. 4-8.
- [7] 7. Rahmatzade, Abolghasem (2008). To where life skills goen? Growth school advisor, educating, analytic, alarming, second period, No.3, P. 12-15.
- [8] 8. Rezaii, Narjes (2003). Study of relation between training of life skills and self-esteem in primary students of Rasht first area, master thesis of common cultural.
- [9] 9. Argyel, M., & Lu, L. (1990) The happiness of extroverts. Personality and individual differenced, 11(1), 10100-1017.
- [10] 10. Carer et al. (2005) Optimism and psychological Health. Positive psychology.
- [11] 11. Cheng Y. Y. Liu. K. S. & Chang. C. M. (2007) The effect of creative problem solving Instruction on elementary schools science lessons (Article written in Chinese) Retrived February 20,2009 From <http://www.fededu.hk/en/cjse/2007150005/0565.cun>.
- [12] 12. Cooper smith (1967) the antecedents of self-esteem free-man and company, Sanfrancisco: P.38.
- [13] 13. Halford, W.K. Markman, H, J, Stanley. S, & Leline, G (2004) Best practice in relationship education. Journal of material and family therapy. 29. (3) 385-406.

Real Time Vehicle License Plate Recognition Based on 2D Haar Discrete Wavelet Transform

R. T. Lee, K. C. Hung, and H. S. Wang

Abstract— This thesis is to present a new approach for license-plate recognition using 2D Haar Discrete Wavelet Transform (HDWT) and Artificial neural Network. This thesis consists of three main parts. The first part is to locate and extract the license-plate in an image. The second part is to train of the license-plate. The third part is to real time scan recognize of the license-plate. Traditional license plate recognition system design complexity. The paper present is a vehicle license plate after 2D Haar Discrete Wavelet Transform three transforms, select only after the third conversion coefficients of low-frequency part of the image pixels, image pixels into one-sixty fourth, thus reducing the number of scanning image pixels, increasing rapid implementation of recognition work and the memory usage. This article is to directly scan for license plate recognition, without recognition of the individual characters. This new approach is a real time recognition, experimental results of license plate recognition rates can be as high as 95.33%.

Index Terms— Haar Discrete Wavelet Transform, real time, Artificial neural Network

1 INTRODUCTION

1.1 The Motive of the Research

TO suggest the that is because of the promotion of life quality, vehicles have become an essential vehicle. The increase of vehicle number produces some management problems, such as the building vehicles, company vehicles management, and so on.

Therefore, we propose that application in a limited district of automatic license plate recognition system to control the restricted districts of the vehicle, it can to save on labor costs and improve efficiency.

1.2 Research of Related Literature

In [1, 3, 4] license plate recognition system are real-time recognition, and in [1, 2, 3, 5, 6, 7] are make a lot of ways the license plate location. In [1] use HSI color space method, the recognition easy light affect. In [2] only make the location of the vehicle's license plate. In [3, 4] are the license-plate characters recognition, but it is not easy to cut character complete and the similar characters confusion problem, such as 1 and I, 0 and O, and so on. In [5] need to find the thresholds of row and column, but affected by light or night and other factors, it can not find the ideal thresholds. In [6] because the image does not at the same position to photography and road has slope. Hence, they will not be effective for the license plate matching. In [7] need to create a reference model, and the standard license plate for recognition, but the position photography is differ-

ent, so its easily lead to recognition errors.

1.3 This Thesis Proposed of Method

Raised in this paper is a license plate recognition and at present Taiwan uses license plate size is 150 mm x 320 mm, license plate have two letters and four-digits consist of each word, words and the interval is 10 mm, lack of space between words and the correct cutting, application of this article to improve the short comings arising out window scanning combination 2D Haar Discrete Wavelet Transformation [2] and Artificial neural Network [11, 12] license plate recognition. To propose the method as following describes.

This articles main license plate recognition system have training and recognition of two parts, training is used vehicle license plate edge functions and Sobel shield edge detection for license plates, and automatically look for vehicle license plate location and extract license plates. License plate recognition have three parts: The first part is a 32x64 resolution window scan, the second part is the use of 2D Haar Discrete Wavelet Transformations, and extract images of low-frequency (LL), the third is to use Artificial neural Network on the limited district of vehicle license plate recognition. Therefore, there character segmentation problems can be avoided.

The advantage of Haar Discrete Wavelet Transform is that it each time transform only needs 1/4 of the original image. Hence, this method can fast execution speed. This paper simulation results by MATLAB software that recognition up to 95.33%.

2 STUDY METHODS

After studying the way is to scan the image in this article, use 2D Haar Discrete Wavelet Transform and Artificial neural Network for recognition. 2.1 chapter is 2D Haar Discrete Wavelet Transform of introduction, 2.2 chapter is the introduction of Artificial neural Network, described later.

R. T. Lee Author is with the Department of Computer and Communication Engineering, National Kaohsiung First University of Science and Technology, Taiwan, R.O.C. (PH-886-986194486 or 886-7-6011000 Ext 2040; fax: 886-7-6429489; e-mail: u9715902@nkfust.edu.tw).

K. C. Hung Author is Department of Computer and Communication Engineering, National Kaohsiung First University of Science and Technology, Taiwan, R.O.C. (PH-886-7-6011000 Ext 2018; fax: 886-7-6011012; e-mail: kchung@nkfust.edu.tw).

H. S. Wang Author is Department of Computer and Communication Engineering, National Kaohsiung First University of Science and Technology, Taiwan, R.O.C. (PH-886-7-6011000 Ext 2020; fax: 886-7-6011012; e-mail: hswang@nkfust.edu.tw).

2.1 2D Haar Discrete Wavelet Transform

The Discrete Wavelet Transform is a very popular method in digital image processing in recent years, especially in multi-resolution representation. The Discrete Wavelet Transform can decompose an image into some sub-bands and encodes the signal of sub-bands according to the importance of signal.

Discrete Wavelet Transform turns an image into high frequency and low frequency data. According to these different data, we can do the processing respectively. Due to the sensitivity of human's eyes, the low frequency data forms the important part of the original image. If the values of low frequency coefficients change, one can recognize the change easily. On the other hand, human's eyes are much less sensitive to the high frequency part, because it is hard for a human to discover it.

The 2D Haar Discrete Wavelet Transform [2] can decompose an image into four sub-bands. Have high frequency (HH) and the high and low frequency (HL) and low and high frequency (LH) and low frequency (LL) the four sub-bands. Low frequency (LL) between the pixels of change is relatively small, image clearer, so after each conversion only to extract low frequency (LL) part. Haar Transform characteristics of: (a) Multiplication is not required. (b) Input and output points the same. (c) Frequency two parts, low frequency all is 1 and high frequency (half are 1, half are -1). (d) You can analyze a signal localized features. (e) Instruction cycle very fast. But it does not suitable for signal analysis.

2.2 Artificial neural Network

Artificial neural Network the network level can be divided into input layer, hidden layer and the output layer and other three. Output transfer function between the layers is to use the hyperbolic tangent function (Tan-sigmoid Transfer function).

Artificial neural Network in this article is the special value output layer neurons converting binary decimal output, and output layer neurons of the components of the target value converted to a decimal value for comparison. Such as output values and target values are the same for training completed, or the express train failed, return train.

3 TRAINING AND RECOGNITION LICENSE PLATE STEPS

In front of the license plate recognition, first through extract license plates of sample use Artificial neural Network training. This license plate recognition system are divided into three parts, the first is a window scan, and the second is the 2D Haar Discrete Wavelet Transform extract images of low-frequency (LL) part is a class of Artificial neural Network recognition. The license plate recognition system is used 32 x 64 resolution windows scans the image, combined with 2D HAAR Discrete Wavelet Transform filter out high frequency part of the image, leaving low-frequency part of the coefficients with using Artificial neural Network to recognition, while the third part of class among the various layers of the Artificial neural Network weights and basis weights are determined by the license plate feature values obtained through the Artificial neural Network training institute.

Following a part of the 3.1 chapter is simulation software

and hardware environment, the 3.2 chapter is vehicle license plate photograph, the 3.3 chapter is license plate training step process instructions, and the 3.4 chapter is license plate recognition step process instructions.

3.1 Simulation Software and Hardware Environment

Using personal computer simulation, the software is Microsoft Windows XP, home edition, version2002, service pack2, hardware is Compaq Presario CQ2020TW computer, Intel@ATOM™ 230 processor, 1024MB DDR2, 160GB hard disk space.

3.2 Vehicle License Plate Photograph

Using the CASIO EXILIM, 10.1 MEGA PIXELS DIGITAL CAMERA EX-S10, adjusting the resolution 480 x 640 for photography vehicle license plates, photograph location diagram as figure 1 of the 12 pieces O sign locations, in the figure indicating the position camera photography, simulation of right side of the forward direction of the cameras installed in vehicles in lanes 25 cm, and camera distance of 300 cm started as photo taken after the first time, each 20 cm according to photo taken at a time until the distance between vehicles 200 cm total photos taken 6; another simulation of vehicle direction if left 50 cm, camera distance 300 cm started as photo taken after the first time, each 20 cm according to photo taken at a time until the distance between vehicles 200 cm total according to photo 6, all according to each vehicle taken 12 when extract license plates as training samples in this article. Forward direction at a distance of vehicles from 200 cm to 300cm of right 25cm and 75 cm (figure 1 have O signs of 12 pieces is photo position) regional routes within each vehicle in any different position in the interregional, according to vehicle license plate location photography like 6, when the license plate recognition sample.

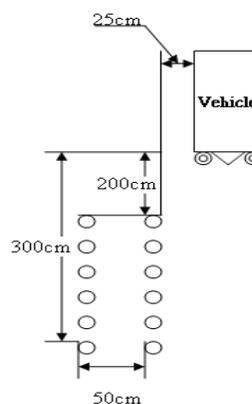


Fig. 1. Vehicle image photo location map

3.3 License Plate Training Step Process

License plate training processes in figure 2 as shown in the license plate train seven steps in this article, described it as follows.

Step1. Enter license plate images, as in figure 5(a).

Step2. Vehicles with license plates Image pre-processing (as figure 5 and figure 7).

2.1). Color photo use (1) equation transform HSI gray-level image, Y are the gray-level image after the conversion and R, G and B are the three elements of the original color image. And because vehicles stepped on induction coil and regional position is fixed, so the image pretreatment, such as crop top 150 pixels, down, left and right around the crop 100 pixel, the crop will not affect the license plate shown, and reduces background noise and run time.

$$Y(i, j) = (R(i, j) + G(i, j) + B(i, j)) / 3 \quad (1)$$

$Y(i, j)$ are gray-scale space red, green and blue color components.

2.2). Vehicle images edge function and Sobel mask operator, and the edge function will automatically detect the critical threshold, the following (2) formula is MATLAB software program. Effects of vehicle images and 0 and 1 binary, making vehicles licensing and background are distinct.

$$\text{Image1} = \text{edge}(\text{Image}, 'sobel') \quad (2)$$

Image for the input image, edge is a instruction, Sobel was operator, Image1 for the output image.

2.3). Using license plate district, height, width, width to height ratio and distance up and down the second one taking a line to extract license plates.

Step3. License plate 2D Haar wavelet Transformation three times, each conversion or last when the license plate features only selected LL low-frequency parts.

Step4. 2D Haar after the discrete wavelet Transform, take the low-frequency part of the data line up. Columns when license plate the feature-values, and the feature coefficients regularization of regularization (3) formula when the artificial neural network input data.

$$\frac{X_i}{\sqrt{\sum_{i=1}^{i=32} X_i}} \quad (3)$$

X_i a line up of train cards coefficients of feature.

Step5. Using Artificial neural Network to train. Set type of Artificial neural Network parameter, enter the number, the number of hidden layer neurons, such as output layer learning rate, number and frequency of training, initial random weight values between layers and partial weight parameter value, and so on. Artificial neural network set the parameters as follows in this article, enter the number of neurons is license plate features factor into a column of values, number of neurons in the hidden layer is 20, number of output layer neurons is 8, and train number 1000, learning rate 0.5, and input layer and hide the layer weight values, hidden and output layers between weight values, input layer with hidden layer basis values, hidden and output layer basis values, the four initial values are randomly generated by a computer program.

Step6. Determine. Up reach to cycle times and convergence. Finally storage weight values and basis values.

Step7. End. After training is complete, get figure 3 training

of vehicle license plate convergence graphs.

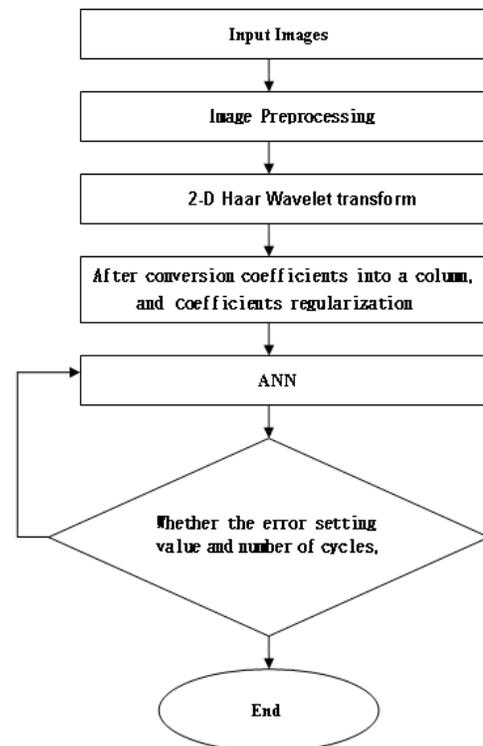


Fig. 2. License plate training flow chart

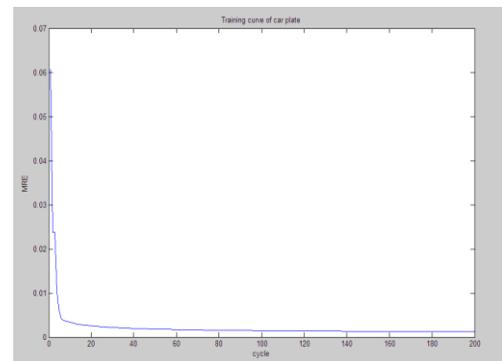


Fig. 3. License plate convergence graph of training

3.4 License Plate Recognition Step Process

License plate recognition system recognition processes such as figure 4. These eight license plate recognition steps are described below:

Step1. Enter recognition plate, such as figure 5(a) first location map, its recognition plate samples such as figure 1, according to 12 photo positions at inter regional, any of 6 different locations, experimental-production 100 vehicles total photos taken 600 sheets when the recognition of samples.

Step2. Vehicles with license plates Image pre-processing (as figure 5 and figure 7).

Step3. 32 x 64 scans the image pixel scan window, and at the same time using 2D Haar discrete wavelet Transform three make the image smaller and only select LL low-frequency parts, in order to reduce the image pixels three times the result

like figure 11. Three pressures after the low frequency parts of vehicle license plate imaging become 60×80 images.

Step4. License plate after 2D Haar Discrete Wavelet Transform, select only the third transformation of LL low-frequency parts.

Step5. The LL low-frequency coefficients into a column, as each scan the regional characteristic of values. After special feature values of regularization, as such Artificial neural Network input data.

Step6. Using Artificial neural Network recognition. Set number input neuron for the license plate number of feature values, and the number of hidden neurons was 20, and the number of output layer neurons was 8, and the load after training is complete Input layer and hide the layer weight values and the hidden and output layer weight values and input layer with hidden layer basis values and the hidden and output layer weights equivalent. Such Artificial neural Network operation output binary strings, and convert the binary string to decimal values.

Step7. Determine. The scanning district feature values, as the Artificial neural Network training algorithm input values, and Artificial neural Network output the correct license plate number, otherwise continue to scan the next district.

Step8. Output the results of recognition.

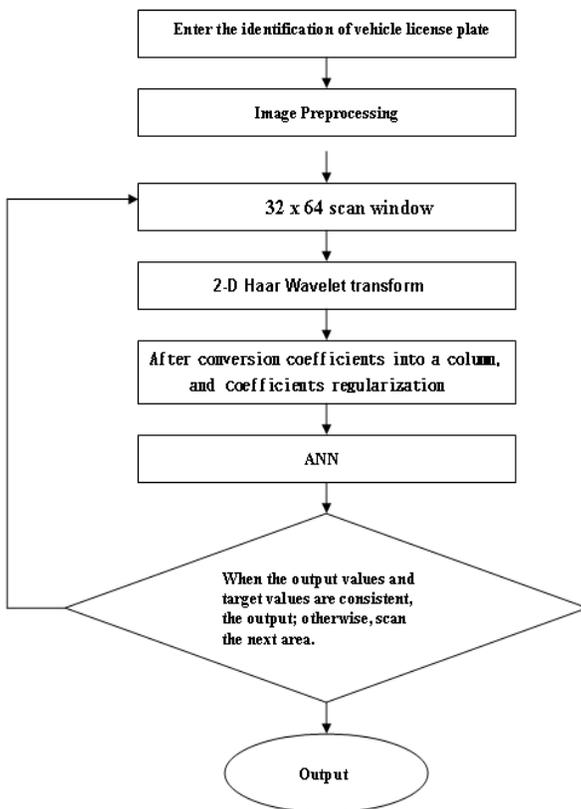


Fig. 4. License plate recognition process.



Fig. 9. (a) During the day license-plate location. (b) License plate



Fig. 10. (a) At night license-plate location. (b) License plate



Fig. 11. (a) 9M-7249 license plate HDWT three times Transform



4 EXPERIMENTAL RESULT AND ANALYSIS

Experimental 1800 cars in the picture, one can see from figure 1 locations per license photo taken 12, a total of 1200 sheets license plate train, while figure 1 locations within each license photo taken 6 pictures, a total of 600 sheets were being used as a test data set, and each image is 24-bit color images. Is in the different location photography, collected the image displays the following properties: (1)The left side photography lead to more complex backgrounds. (2)Some pictures as the evening sunlight and low quality. (3)According to the distance was taken different plate sizes vary.

This section describes the 4.1 chapter is a license plate recognition results figure and table, the 4.2 chapter is the recognition and analysis of the results. This article in license plate training step first using edge function and Sobel shield operator for edge reconnaissance pre-processing, and automatically looking for license plate location and the extract license plate, in 1200 sheets image have 80 sheets license plate location failed, failed causes have car lamp and too sun strongly reflection or reconnaissance to similar license plate background, and so on problems, extract license plates accuracy rate of only about 93.3%, such as table 1 shows, training and license plate number, such as table 2 shows; extract license plates used for recognition of the correct, by table 3 license plate recognition result that recognition rate of 95.33%, so the license plate recognition system is the most values.

4.1 License Plate Recognition Results Figure And Table

License plate recognition results as figure 12 and figure 13 shows.



Fig. 12. During the day license plate recognition result.



Fig. 13. At night license plate recognition result.

TABLE 2. LICENSE PLATE NUMBER TRAINING TABLES

Number of per vehicle training	per vehicle training 1 sheet	per vehicle training 12 sheets
Train number license plate (sheets)	100	1200

TABLE 3. LICENSE PLATE RECOGNITION RESULT TABLE

Per vehicle recognition number	per vehicle recognition 1	per vehicle recognition 6
Vehicle recognition number (sheets)	100	600
Recognition number of successful (sheets)	93	572
Recognition number failed (sheets)	7	28
Recognition rates (%)	93	95.33

4.2 Recognition and Analysis of Results

This image is covered by the limited number of vehicles and vehicle license plates in a certain region of space, and the day at the normal climate as taken in the evening. Recognition of vehicle license plate was shooting at day and night mixing and a weather well, and vehicle license plate location anywhere within the region 6, and 100 vehicles, all total 600 samples as license plate recognition. For each image 32×64 pixels scan window to scan, and three times 2-D Haar Discrete Wavelet transform, filter out high frequency part of the pixel, only extract third Haar Discrete Wavelet Transform of low-frequency parts of coefficients with the image into 4×8 , this method reduces the image factor and fast recognition speed of execution. And load training network weight values and and basis weights, after we use Artificial neural Network to recognition of the vehicle of a district, and finally output the vehicle recognition results.

This license plate recognition system first steps is using edge function and Sobel shield operator for license plate edge reconnaissance, and automatically looking for license plate location and the extract license plate, we will extract license plate through class Artificial neural Network training, to obtained entered layer and hidden layer weight values and the hidden layer and output layer weight values, and the entered layer and hidden layer partial right value and the hidden layer

TABLE 1. LICENSE PLATE CAPTURE RESULTS TABLE

The number of per vehicle	per vehicle training 12
---------------------------	-------------------------

and output layer partial right value, parameter value. Step two is to use 32×64 pixels scan window is different from the training samples of unknown license plate image scanning, and through 2-D Haar Discrete Wavelet Transform three times and take the third transformation of low-frequency coefficient and load weight values of the first steps training and partial weight and other parameters and their values are fixed, unchanging, and in a limited district using Artificial neural Network to recognition of vehicles. Simply speaking, to use the unknown license plate feature values to be kind of Artificial neural Network operations, such as match unknown license plate number is one of the original training sample license plates. Finally, recognition results of output, if the recognition yes, opening the gates for a vehicle to enter. Otherwise recognition fail, prohibition of vehicular enter.

This article by using MATLAB software laboratory test results, the vehicles according to 600 sheets have 28 sheets failed, his resolution of 95.33%.

5 CONCLUSION

This papers by made of license plate recognition system is different from traditional of license plate recognition system, its characteristics is directly scan and the using 2-D Haar Discrete Wavelet three times conversion, and every time conversion image take low frequency part, makes image became original image of $1/4$ coefficients, and by three times conversion image extract low frequency part of vehicles license plate image, make image coefficients became original of one-sixty fourth, thus reducing coefficients of recognition plate imaging, it can fast recognition speed.

This paper Recognition system are combination of 2-D Haar Discrete Wavelet transform and artificial neural networks are license plates through the 2-D Haar Discrete Wavelet transform, get the license plate of the feature values, and characteristic values into a column, then the feature values corresponding to license plate type Artificial neural Network training, when training is complete, it using training received weight values load to Artificial neural Networks. Then, using 32×64 scan window progressive scan unknown effects of vehicle license plate recognition, and outputs the results. MATLAB software test result, license plate recognition rate up to 95.33%.

REFERENCES

- [1] H. Sheng, C. Li, Q. Wen, and Z. Xiong, "Real-Time Anti-Interference Location of Vehicle License Plates Using High-Definition Video," *IEEE Intelligent Transportation Systems Magazine*, 2009.
- [2] M.-K. Wu, L.-S. Wei, H.-C. Shih, and C. C. Ho#, "License Plate Detection Based on 2-Level 1-D Discrete Periodic Wavelet Transform and Edge Density Verification," Embedded SoC Lab, Department of Electrical Engineering, National Yunlin University of Science and Technology, Douliou, Yunlin 64002, Taiwan, ROC, *IEEE International Symposium on Industrial Electronics (ISIE 2009)*, Seoul Olympic Parktel, Seoul, Korea, July 5-8, 2009.
- [3] A. Conci., J. E. R. d. Carvalho and T. W. Rauber, "A Complete System for Vehicle Plate Localization, Segmentation and Recognition in Real Life Scene," *IEEE Latin America Transactions*, Vol. 7, No.5, September

- 2009.
- [4] L. Hai, and G. Shrestha, "Real Time License Plate Recognition at the Time of Red Light Violation," College of Automation, Harbin Engineering University, Harbin, China, *International Conference on Computer Application and System Modeling (ICASM 2010)*, 2010.
- [5] P. Kulkarni, A. Khatri, P. Banga and K. Shah, "A Feature Based Approach for Localization of Indian Number Plates," Dept. of Electronics and elecommunication, Smt. Kashibai Navale College of Engineering, University of Pune, India, *IEEE*, 2009.
- [6] Y. Song, "A Level Set Based Method of License Plate Localization," Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, *Chinese Academy of Sciences, IEEE IRI 2010*, Las Vegas, Nevada, USA, August 4-6, IEEE, 2010.
- [7] I. Daubechies, "Orthonormal Based of Compactly SupportWavelet," *Communications in Pure Applied Math*, vol.41, pp.909-996, 1988.
- [8] K. C. Hung, Y. J. Huang, F. C. Hsieh and J. C. Wang, "An AOCA-based VLSI architecture for non-recursive 2D discrete periodized wavelet transform," *Digital Signal Processing*, 2002. 2002 14th International Conference on Volume 1, vol.1, pp.273-276, July 2002.
- [9] C. F. Tsai, H. S. Wang, and K. C. Hung, "A high efficient non-recursive discrete periodized wavelet transform for extracting the transformed coefficients of coarser resolution levels," *The 2004 IEEE Asia-Pacific Conf. on Circuits And Systems*, pp. 661-664, Dec. 2004.
- [10] A. S. Lewis and G. Knowles, "VLSI architecture for 2D Daubechies wavelet transform without multipliers," *Electronics Letters*, vol. 27, no. 2, pp. 171-173, Jan. 1991.
- [11] P. M P, M. S. A.I Majid, S. Yaacob, M. H. F. Rahiman and R. P. Krishnan, "Statistical Time Energy Based Damage Detection in Steel Plates Using Artificial neural Networks," School of Mechatronics Engineering, University Malaysia Perlis, Perlis, Malaysia, *International Colloquium on Signal Processing & Its Applications (CSPA)*, IEEE, May, 2009.
- [12] N. Chen and L. Xing, "Research of License Plate Recognition based on Improved BP Artificial neural Network," College of computer science Xi'an Polytechnic University Xi'an, China, *International Conference on Computer Application and System Modeling*, 2010.

Bai's Secret Sharing Scheme for Highly Secured AOMDV Routing Protocol

C. Chandrasekar, *Lt.Dr. S. Santhosh Baboo*

Abstract--- Security in wireless network is of a great concern in the present wireless environment. Security is becoming a much essential part in wireless communication as intruders are very much keen in attacking the transmitted route or any node to steal the data packets. It is indispensable to provide security to the network from the intruders and their security attacks. In order to provide a secured environment, it is essential that routing protocol exploited in a MANET should be of highly secured from the intruders. There are several routing protocols exist but most of these protocols makes use of only a single path. Therefore, proposed a Highly Secured Ad-hoc On-demand Multipath Distance Vector routing protocol (HSAOMDV) using the Bai's secret sharing scheme. Secret sharing scheme allows to distribute a secret only between a set of participants in a communication process by providing each participant a share. By this manner only the group of participants are allowed to reconstruct the secret. In order to evaluate the performance of the proposed HSAOMDV, it is compared against AOMDV and SAOMDV (AOMDV with Shamir's Secret Sharing Scheme). Number of packets received by HSAOMDV routing protocol does not decreased with intrusion of malicious nodes in the network and the packet delivery ratio also comparably high than the other two routing protocols.

Keywords--- Mobile Adhoc Networks, Multihop Wireless Networks, Ad-hoc On-Demand Multipath Distance Vector (AOMDV), Bai's Secret Sharing Scheme

I. INTRODUCTION

Mobile Ad-hoc NETWORK (MANET) is a mobile, multihop wireless network that does not depend on any predefined infrastructure. MANETs are featured by active topologies because of their uncontrolled node mobility, inadequate and uneven shared wireless channel bandwidth and wireless devices constrained by battery power. The major challenge in MANETs is to intend dynamic routing protocols that are proficient with very less overhead.

Security in MANETs is of a major concern and it has become an active area of research. To prevent a variety of attacks in MANETs has been a challenging issue for researchers as MANETs has been widely used in military applications, emergency rescue operations, in confidential video conferencing, etc. A MANET is an automatic network which is fully active and spread in nature. The operations of every node are similar but the recognition of an attacker or malevolent (malicious) nodes amongst the network is a difficult task. Recently, the security for multicast routing in MANETs has also become very vital. Several security protocols have been developed under the operations of multicast [1]. However, these protocols are susceptible to several types of attacks on MANETs [2] like flooding, blackhole, wormhole, etc. Various researches are being done for handling the attacks in MANETs.

The major purpose of using routing protocols in an ad-hoc network is to facilitate the source to identify routes to destination with the cooperation of other nodes. Because of the random movement of the nodes, the network topology alters quickly and arbitrarily. Thus, the routing protocol must be capable of handling these alterations and must facilitate the nodes to find new routes to sustain connectivity.

The security in MANETs [3][4] has become a serious issue mainly because of the active characteristic of the ad hoc network and because of the necessity to function efficiently with inadequate resources, including network bandwidth and the CPU processing capacity, memory and battery power (energy) of each individual node in the network. Quick and frequent routing protocol communication between nodes is very much needed [13].

In this paper, Bai's secret sharing scheme is used for the purpose of providing better security among MANET nodes. The secret key sharing scheme is used in this approach as it provides assurance to the source node or the owner regarding the genuinely participating nodes in the network. The main aim is to sustain the genuineness of the nodes available in the network.

Ad-hoc On-Demand Multipath Distance Vector (AOMDV) is one of the potential routing protocols for maintaining security. Several intruders are very much interested in attacking the nodes or route to steal the data packets [15]. In order to provide more security, Bai's secret sharing scheme is incorporated to AOMDV to make it as Highly Secured AOMDV (HSAOMDV). In HSAOMDV, only the authorized users are allowed to take part in the communication.

II. LITERATURE SURVEY

MANETs has several kinds of security issues, caused by their nature of collaborative and open systems and by limited availability of resources. In this paper, Cerri et al., [5] consider a Wi-Fi connectivity data link layer as a fundamental technique and concentrates on routing security. The author discusses the implementation of the secure AODV protocol extension, which comprises of alteration policies aimed at enhancing its performance [14]. The author proposed an adaptive technique that adjusts SAODV behaviour. Furthermore, the author examined the adaptive technique and another approach that delays the verification of digital signatures. This paper sums up the experimental results collected in the prototype design, implementation, and tuning.

Li Bai et al., [6] proposed a strong (k,n) threshold-based ramp secret sharing scheme with k access levels. The secrets are the elements represented in a square matrix. The secret matrix S can be shared with n different users by means of a matrix projection technique in which: (a) any subset of k users can work together to reconstruct the secret and (b) any subset of $(k - 1)$ or fewer users cannot partly identify the secret matrix. The essential benefits are its large compression rate on the size of the shares and its strong protection of the secrets.

C. Chandrasekar, *Research Scholar*, Computer Science, Manonmaniam Sundaranar University, Tirunelveli - 627 012, chandrasekar2000@gmail.com

Lt.Dr. S. Santhosh Baboo, *Reader, P.G. and Research*, Dept of Computer Science, D.G.Vaishnav College, Chennai - 600106.

Perlman proposed a link state routing protocol [7] that attains Byzantine strength. Though, the protocol is extremely forceful, it needs a very high operating cost associated with public key encryption. Zhou and Haas [8] chiefly describe key management in their paper to provide security to ad hoc networks. The author devotes a part to secure routing, but in essence concludes that “nodes can defend routing data in the similar way they protect data traffic”. They also examine that denial-of-service attacks against routing will be considered as damage and it is routed around. Certain research has been done to secure ad hoc networks by means of misbehaviour detection approaches. This technique has two major problems: Initially, it is fairly likely that it will be not possible to discover various kinds of misbehaving; and secondly, it has no real means to assure the integrity and authentication of the routing messages.

Multipath routing diminishes the penalty of security attacks obtaining from collaborating malevolent nodes in MANET, by increasing the number of nodes that an opponent must negotiate in order to take control of the communication. In this paper, various attacks that cause multipath routing protocols more susceptible to attacks than it is expected, to collaborating malevolent nodes are recognized. Kotzanikolaou et al., [9] proposed a novel On-demand Multipath routing protocol called the Secure Multipath Routing protocol (SecMR) and the author examine its security properties. The SecMR protocol can be easily combined in an extensive variety of on-demand routing protocols, such as DSR and AODV.

III. METHODOLOGY

3.1. Ad hoc On-Demand Multipath Distance Vector Routing

The significant purpose of this paper is to provide a highly secured AOMDV routing protocol by incorporating Bai’s secret sharing scheme. In this section the goal is to enhance the AOMDV protocol to work out multiple disjoint loop-free paths in a route discovery. AOMDV can be implemented even in the existence of unidirectional links with other techniques to assist in discovering bidirectional paths in such circumstances [10].

AOMDV has numerous features which are similar with AODV. It is dependent on the distance vector theory and utilizes hop-by-hop routing technique. Furthermore, AOMDV also discovers routes on demand using a route discovery method. The most important variation is the amount of routes found in each route discovery. In AOMDV, RREQ transmission from the source to the target establishes multiple reverse paths both at intermediary nodes in addition to the destination. Multiple RREPs navigates this reverse route back to form multiple onward routes to the target at the source and intermediary nodes. Moreover, AOMDV also makes intermediary nodes available with alternate routes since they are established to be helpful in dropping route discovery frequency [11].

The basis of the AOMDV protocol lies in guarantee that multiple routes revealed are loop-free and disjoint, and in competently discovering such paths by means of a flood-based route discovery. AOMDV path revise rules, exploited locally at every node, play a major role in preserving loop-freedom and disjointness characteristics.

AOMDV depends more on the routing information previously available in the fundamental AODV protocol, thus preventing the overhead acquired in determining multiple paths. Specifically, it does not make use of any particular control packets. Additional RREPs and RERRs for multipath discovery and protection together with a small amount of extra fields in routing control packets (i.e., RREQs, RREPs, and RERRs) comprise the only extra overhead in AOMDV compared with AODV.

Disjoint Paths

In addition with continuing multiple loop-free paths, AOMDV looks for to discovering disjoint alternate routes. In order to improve the fault tolerance by means of multiple paths, disjoint paths are an essential alternative for choosing an efficient subset of alternative routes from a potentially huge set since the probability of their associated and simultaneous failure is less important when compared to overlapping alternate routes. Two categories of disjoint paths are taken into account: link disjoint and node disjoint. Link disjoint is a set of routes between a pair of nodes which does not have any mutual links, while node-disjointness in addition prevents mutual intermediary nodes.

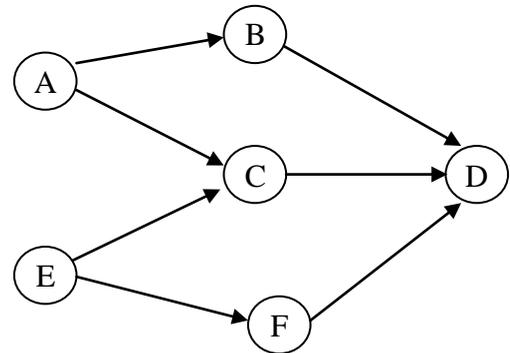


Figure 3.1: Paths maintained at different nodes to a destination possibly will not be equally disjoint.

Here D is the target. Node A has two disjoint paths to D: A - B - D and A - C - D. In the same way, node E has two disjoint paths to D: E - C - D and E - F - D. However the paths A - C - D and E - C - D are not disjoint; they share a mutual link C - D.

3.2. Secret Sharing in AOMDV

SAOMDV with Shamir’s secret sharing scheme is described in short before introducing the Bai’s secret sharing scheme using matrix projection.

SAOMDV with Shamir’s Secret Sharing Scheme

Shamir [12] developed the scheme of a (k, n) threshold-dependent secret sharing scheme. This scheme permits a polynomial function of the $(k - 1)$ th power as,

$$f(x) = s + d_1x + d_2x^2 + \dots + d_{k-1}x^{k-1} \pmod{p},$$

in which the value s is the secret, d_i are arbitrarily chosen values and p is a prime number. The secret shares are the pairs of values (x_i, y_i) where $y_i = f(x_i), 1 \leq i \leq n$ and $0 < x_1 < x_2 \dots < x_n \leq p - 1$.

The polynomial function $f(x)$ is destroyed after each shareholder possesses a pair of values (x_i, y_i) in order that no single shareholder recognizes the secret value s . No groups of $(k - 1)$ or smaller amount secret shares can find out the secret s . Alternatively, when k or more secret shares are accessible, then might set up at least k linear equations $y_i = f(x_i)$ for the unidentified x_i ’s and s . The distinctive solution to these equations confirms that the secret value s can be easily acquired by means of Lagrange interpolation.

Major Disadvantages of using Shamir’s Secret Sharing Scheme in AOMDV

- One of the major concerns of this scheme is that it has an extremely easy access structure. Any k out of n participants are capable of reconstructing the secret.
- An additional disadvantage of this scheme is that if the size of the share enlarges with the increase in the number of participants in a communication process, the size of the share also becomes larger.
- The most common disadvantage of this Shamir’s secret sharing scheme is its complexity. In case

of exploiting a secret sharing scheme in a wireless networking environment, it must be noted that it should not generate intolerable delays.

3.3 Bai's Secret Sharing Scheme using Matrix Projection

In this section designed highly secured AOMDV with Bai's secret sharing scheme.

Consider A be an $m \times k$ matrix with rank $k(m \geq k > 0)$, and

$$S = A(A'A)^{-1}A'$$

where $(\cdot)'$ is the transpose of a matrix and $m \times m$ matrix S is A's projection matrix,

$$S = \text{proj}(A)$$

The vectors v_i using k can be calculated linearly independent $k \times 1$ vectors x_i ,

$$v_i = Ax_i,$$

in which $1 \leq i \leq k$. These $m \times 1$ vectors v_i can be placed in

$$B = [v_1 \ v_2 \ \dots \ v_k].$$

Based on the Invariance Theorem [3], the projection matrix of B must be similar as that of A:

$$A(A'A)^{-1}A' = B(B'B)^{-1}B'$$

3.3.1. Steps in Bai's Secret Sharing Scheme

Following is the procedure used for the secret sharing in HSAOMDV.

Phase I : Construction of Shares

For a secret matrix S, the shares are generated as:

1. Create an $m \times k$ random matrix A of rank k in which $m > 2k - 3$,
2. Select n random $k \times 1$ vectors x_i (any k vectors are linearly independent),
3. Compute share $v_i = (A \times x_i) \pmod p$ for $1 \leq i \leq n$.
4. Calculate $S = \text{proj}(A)$,
Calculate $R = (s - S) \pmod p$,
5. Destroy matrix A, x_i s, and S and s.
6. Distribute n shares v_i and make R publicly known.

Phase II : Reconstruction of Secret

In order to reconstruct, the secret s with k or more shares v_i , the procedure is followed,

7. Collect shares $v_{i_1} \ v_{i_2} \ \dots \ v_{i_k}$
8. Generate a matrix B using only k shares as $B = [v_{i_1} \ v_{i_2} \ \dots \ v_{i_k}]$.
9. Compute the projection matrix $S = \text{proj}(B)$,
10. Calculate $S = (S + R) \pmod p$.

It is to be observed that in step 1 of phase one, the constraint that $m > 2k - 3$ comes from a condition to appropriately apply matrix projection technique to multiple secret sharing.

In steps 2 and 3 of phase two, once the groups of participants have calculated B and S, they must make sure whether B has a rank k. If any of these steps fail to satisfy, it is identified that there must be some unintentional errors or dishonest behavior or intruders attack. By this manner, this scheme also becomes partially verifiable.

IV . EXPERIMENTAL RESULTS

A simulation testbed for MANET is built up to evaluate the performance of the proposed HSAOMDV against AOMDV and SAOMDV routing protocol. All these protocols were experimented over this testbed and its performance was evaluated based on different scenarios.

The values of some constraints considered during the

evaluation are noted below.

Area	1500*300 meter ²
One time quantum	50 msec
Speed of the nodes	20 meters/second
Run time for the simulation	200 second
Direct Transmission Range of the nodes	250 meter
Channel capacity	Mbps

4.1. No. of Data Packets Vs No. of Malicious Nodes

In order to find the effect of malicious nodes in all these routing protocols, initially 7000 data packets are sent. From the figure 4.1, it is revealed that with the increase in the quantity of malicious nodes, the number of data packets received by AOMDV and SAOMDV decreases considerably, but the number of data packets received by HSAOMDV remains almost steady. It indicates that malicious nodes have only lesser consequence on the amount of data packets transmitted by HSAOMDV. Simultaneously, the data packets received in AOMDV and SAOMDV falls considerably with the increase in the number of malicious nodes.

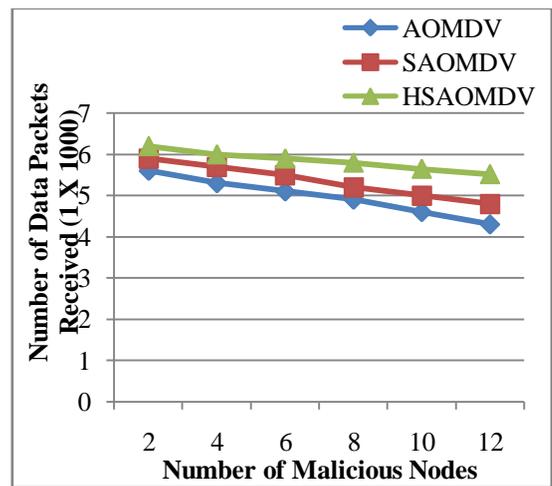


Figure 4.1: No. of Data Packets Vs No. of Malicious Nodes

4.2 Packet Delivery Ratio (PDR) Vs No. of Malicious Nodes

Packet Delivery Ratio (PDR) is the proportion of the amount of data packets received by the target node to the amount of data packets transmitted by the source node. It is clear from figure 4.2 that PDR of AOMDV and SAOMDV is significantly affected by the introduction malicious nodes while the PDR of HSAOMDV is unaffected to it.

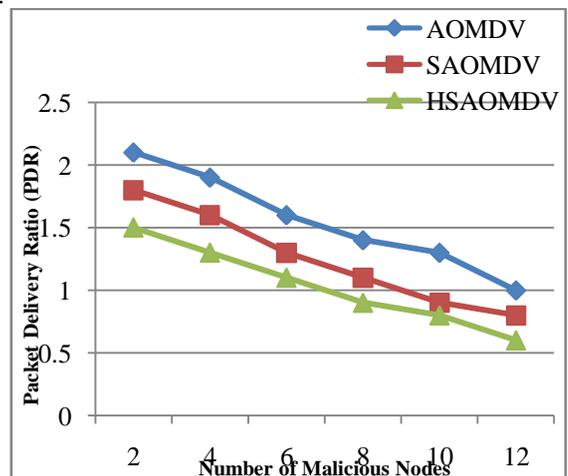


Figure 4.2: Packet Delivery Ratio (PDR) Vs No. of Malicious Nodes

V. CONCLUSION

Security of MANET is becoming a major concern and it is turning out to be an essential research in the field of wireless networks. A highly secured routing protocol must provide security against the intruders and the attacks. The proposed routing protocol combines the AOMDV with Bai's secret sharing scheme with the purpose of providing a secured communication. In this paper, a highly secured ad-hoc on-demand multipath distance vector routing protocol called HSAOMDV. This scheme helps in easily recognizing the malicious nodes by providing proper authorization since only the users with proper key are allowed to share and reconstruct the secret. The simulation result confirms that HSAOMDV is less vulnerable to the intrusion of malicious nodes and moreover there is no effect in the packet delivery ratio.

References

- [1] Luo Junhai, Ye Danxia, Xue Liu, and Fan Mingyu, "A Survey of Multicast Routing Protocols for Mobile Ad-Hoc Networks", IEEE Communications Surveys & Tutorials, Vol. 11, No. 1, Pp. 78-91, 2009.
- [2] Bounpadith Kannhavong, Hidehisa Nakayama, Yoshiaki Nemoto, Nei Kato, "Survey of routing attacks in Mobile Ad-Hoc Networks", IEEE wireless communication, Pp. 85-91, 2007.
- [3] Hongmei Deng, Wei Li, and Dharma P. Agarwal "Routing security in wireless Ad Hoc networks", IEEE Communications Magazine, 2002.
- [4] M. Guerrero Zapata and N. Asokan, "Securing Ad hoc Routing Protocols," in Proceedings of the 1st ACM workshop on Wireless security, Atlanta, GA, USA, Sep 2002, pp. 1-10.
- [5] Cerri, D. Ghioni, A. "Securing AODV: the A-SAODV secure routing prototype", IEEE Communications Magazine, Vol. 46, No. 2, page(s): 120 - 125, 2008.
- [6] Li Bai, "A Strong Ramp Secret Sharing Scheme Using Matrix Projection", Proceedings of International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'06), Pp. 656, 2006.
- [7] R. Perlman, Fault-tolerant broadcast of routing information, Computer Networks, 7, 395-405 (1983).
- [8] L. Zhou, and Z. J. Haas, Securing ad-hoc networks, IEEE Network Mag., 13, 24-30 (1999).
- [9] Kotzanikolaou, P.; Mavropodi, R.; Douligeris, C.; "Secure Multipath Routing for Mobile Ad Hoc Networks", WONS 2005. Second Annual Conference on Wireless On-demand Network Systems and Services, Page(s): 89 - 96, 2005.
- [10] Marina MK, Das SR, "Routing performance in the presence of unidirectional links in multihop wireless networks", In Proceedings of ACM MobiHoc, 2002.
- [11] Nasipuri A, Castaneda R, Das SR, "Performance of multipath routing for on-demand protocols in mobile ad hoc networks", ACM/Kluwer Mobile Networks and Applications (MONET), Vol. 6, No. 4, Pp. 339-349, 2001.
- [12] Adi Shamir, "How to Share a Secret", Communications of the ACM, Vol. 22, No. 11, Pp. 612-613, 1979.
- [13] Marina MK, Das SR, "Performance of route caching strategies in dynamic source routing", In Proceedings of Workshop on Wireless Networks and Mobile Computing (WNMC) in conjunction with International Conference on Distributed Computing Systems (ICDCS), 2001.
- [14] Perkins CE, Belding-Royer E, Das SR, "Ad hoc on-demand distance vector (AODV) routing", <http://www.ietf.org/rfc/rfc3561.txt>, 2003.
- [15] Kumar, H.; Sarma, D.; Kar, A.; "Security threats in wireless sensor networks", IEEE Aerospace and Electronic Systems Magazine, Vol. 23, No. 6, Pp. 39 - 45, 2008.

Factors Affecting the Performance of Hindi Language searching on web: An Experimental Study.

Kumar Sourabh Vibhakar Mansotra

Abstract

With the internet growing at an exponential rate the web is increasingly hosting web pages in different languages. It is essential for the search engines to be able to search information stored in a specific language. The native users also tend to look for any information on web nowadays. This leads to the need of effective search engines to fulfill native user's needs and provide them information in their native languages. The major population of India use Hindi as a first language. The Indian constitution identifies 22 languages, of which six languages (Hindi, Telugu, Tamil, Bengali, Marathi and Gujarati) are spoken by at least 50 million people within the boundaries of the country—there are a large number of them living outside the country. The Hindi language web information retrieval is not in a satisfactory condition. The presence of Hindi on the World Wide Web is still limited and tentative because of attitudinal and technical factors. Besides the other technical setbacks the Hindi language search engines face the problem of morphology, phonetics, word sense disambiguation etc. The performance of search engines is affected by these problems. This paper covers the comprehensive analysis and also the comparison of the affect of language structure related factors (morphology, phonetics, WSD, synonyms,) on the performance of search engines supporting Hindi language.

Keywords: search engines, morphology, word sense disambiguation, precision, Guruji, Raftaar and Hindi Language.

1. Introduction

With the web content being written in different languages of the world, it has become important to have tools that can retrieve information from the documents written in different languages. In the context of Indian languages, Hindi language has been given much emphasis leading to the development of significant number of Hindi documents. In fact, of the top 100 languages in the world, English occupies the top position, with Hindi coming fifth. [1].

Hindi language information retrieval on the web is still in its nascent stage. The number

of users who want the information in Hindi language is increasing. This leads to the demand of the Hindi information retrieval on the web. It is the fact that to date Internet is vigorously used in India by the people who are comfortable in English language. The under development of web in Indian regional languages is one of the important reasons behind the limited growth of Internet in India. Indians use 22 official languages and 11 written script forms and among all the languages Hindi language is spoken by the major population of India. About 5% of population understands English as their second language. Hindi is spoken about 30% of the population [2].

It is the language of dozens of major newspapers, magazines, radio and television stations and of

-
- *Kumar Sourabh is currently pursuing PhD degree program in Computer science University of Jammu, India, PH-9469163570. E-mail: kumar9211.sourabhe@gmail.com*
 - *Vibhakar Mansotra is currently Associate professor in University of Jammu, India, PH-9419103488. E-mail:vibhakar20@yahoo.co.in*

other media. This generates the need of the development of the powerful tools for Hindi language information retrieval. [3].

2. Encoding Standards for Indian Languages

Majority of information seekers use a search engine to begin their Web activity. In this case, users submit a query, typically a list of keywords, and receive a list of Web pages that may be relevant, typically pages that contain the keywords. Today though considerable amount of content is available in Indian languages, users are unable to search such content because Indian language websites rely on unique encodings or proprietary extensions of existing standard encodings [4]

The two main standards in character representation of Indian languages are ISCII and Unicode.

2.1 Indian Standard Code for Information Interchange (ISCII).

2.2 Unicode

The Unicode standard provides with three encoding formats: UTF-8, UTF-16 and UTF-32. Any one of these forms can be used to represent the Unicode characters. Each of these is used in different environments. The default encoding form of Unicode is UTF-16. [5]

Most often it has been observed that the use of proprietary fonts of different standards in Central Government Offices in India, which are not compatible with each other, is causing serious problems in information exchange amongst these offices, In order to facilitate free exchange of information/files/documents. Department of Information Technology, Government of India has accepted Unicode encoding for fonts as Indian standard in this regard. [6]

3. Hindi language and web searching

The spread of Internet in India is today constrained by the fact that mostly the English knowing have been benefited by Internet which is a disappointing situation as the real benefit of

internet does not reach to the common man having less/no knowledge of English language.

The under development of Web in Indian regional languages is one of the important reasons behind the limited growth of Internet in India. A recent survey by a Delhi based research organization - Juxt Consult - says that 44 % of existing Internet users in India prefer Hindi to English, if made available. Similarly 25% existing Internet users prefer other regional languages. Many big companies like Google, Yahoo and Sify are also taking big steps in Hindi and other regional languages. Despite the latent demand among Internet users for Hindi, if there is very dismal use of Hindi, it is due to certain constraints. These include technological, attitudinal and economic factors. The most important hardware used for internet surfing is the keyboard. Various Hindi keyboards are available in different varieties. Most of the keyboards are phonetically different. However, a detailed analysis of whether these are truly optimal or better arrangements exist, has not been done. Most of this research has been in two broad directions: *Normal Keyboards* *Ambiguous Keyboards* [7]

Another constraint in spread of Hindi over the WEB is that of limited content. Where there are more than 20 billion pages on web in English, this number is not more than 10 million in Hindi. This poverty of content is partly due to technological factors and partly due to attitudinal. It is a big dilemma that on the one hand the number of Hindi readers and number of Hindi-speaking people using mobile and computers is so large, and on the other hand the websites are very limited in their content and number.[8] This dilemma should be overcome as soon as possible.

4. Search engines supporting Hindi Language

Now a day's various search engines support information retrieval in different languages. Google, yahoo, Bing, AltaVista are popular worldwide for searching the web. Recently Hindi search engines like Guruji and Raftaar from India have emerged out for information retrieval in Hindi Language. These Indian search engines are new as compared to international search engines

listed above. These search engines find their usage in India for Hindi IR. In this Paper we chose Google, Bing and Guruji for experiment purposes based upon their usage and popularity in India.

5. Factors affecting performance of Hindi language searching on web.

Factors which affect searching Hindi information on web are.

- **Morphological Factors:** Morphology is the branch of linguistics that studies patterns of word formation within and across languages, and attempts to formulate rules that model the knowledge of the speakers of those languages. [9]
- **Phonetic nature of Hindi Language:** Many different languages are spoken in India, each language being the mother tongue of tens of millions of people. While the languages and scripts are distinct from each other, the grammar and the alphabet are similar to a large extent. One common feature is that all the Indian languages are phonetic in nature. [10]
- **Words Synonyms:** India has rich diversity in languages, culture, customs and religions. But, the language structure and variation in dialects is making hindrances in the advantages of Information retrieval revolution in India. For example: we know God is named as "भगवान" in Hindi but we can also call "भगवान" as "प्रभु" इश्वर" or "देवता" and more. It is difficult to decide that which one is to choose?
- **Ambiguous Words:** Many words are polysemous in nature. Finding the correct sense of the words in a given context is an intricate task. One word has more than one meaning and meaning of word is depends on context of sentence. Example

कर (Tax) having synonyms ब्याज, शुल्क, सूद, महसूल, टैक्स in one context and in another context कर (Hand or arms) हस्त, बाँह, आच, शबर and कर (to do) करना in another context.

In this paper we have focused on these four major and critical problems, details and experimental analysis are discussed below.

6 Experimental Study On

6.1 Morphological Factors

Hindi language is morphologically rich language. It has well defined morphological structure and well defined grammar. But the grammatical and language structural standard is least followed due to various reasons. One of the reasons is the language diversity in India. Including Hindi there are about 28 Languages spoken in India and Hindi being the National Language of India is influenced by the regional languages which results a change in dialects not only in speaking but writing also. Every language uses some markers like (English language uses s, es, ing and ऐं, यां, े, औं MAATRAAS in hindi language) are used with a root word and new words are constructed . For ex. (Planning in English) *Yojnaaon* योजनाओं, *Yojnaayein* योजनाएँ in Hindi, are the morphological variants of root word *Yojnaa* योजना. It is desirable to combine all the morphological variants of the words in a single canonical form. The process is called as word stemming and this canonical form is called as root word or base word. We have taken a sample set of 50 queries to test the affect of the root word. Following table (6.1) is the set of randomly selected queries from the set which throw light on the effect of the root word on the performance of Hindi language search engines. Table 6.1.1 shows the results of experiments for effect of morphological factors on Hindi queries.

(Space for Tables 6.1 and 6.1.1)

It has been observed that documents returned by all three search engines are more in number when query with root word is submitted. This justifies the searching of documents in the root word

because in general we get better results with the keywords in their root form.

It has also been observed that only Google shows listing of morphological variants of root words, where as Bing and Guruji show only listing of root word supplied in almost all the sample queries listed above in the table.

From the above results it is evident that only Google indexes the documents keyword in their root form. Bing and Guruji do not index in that form that is the reason number of documents retrieved in their case is less in comparison to Google. The overall comparison of results from the three search engines in tables above show that in general the quantity of results retrieved increased when the keywords are used in their root form. In case of search engines the quality of results is more important than the quantity. Table 6.1.2 and Graph (6.1.3) shows the comparison of the precision values of the three search engines. The precision value is calculated by taking the top 10 results of the search engines. On closely observing the results we can say that precision value in case of Google is high in almost all queries. As mentioned above Google does its indexing in the root form of keywords it can be concluded from the table above that relevancy of the results is also high in Google in comparison to other two search engines which denotes that not only quantity but the quality of results is also affected by the morphological variations in the keywords.

6.2 Phonetic nature of Hindi Language and Spelling variations

The major reasons for spelling variations in language can be attributed to the phonetic nature of Indian languages and multiple dialects, transliteration of proper names, words borrowed from regional and foreign languages, and the phonetic variety in Indian language alphabet. The variety in the alphabet, different dialects and influence of regional and foreign languages has resulted in spelling variations of the same word [11]. For example; Following are the possible spelling variations for the Hindi word अंग्रेजी (angrējī):

(means English)
अंग्रेजी, अंगरेजी, अन्ग्रेजी, अंगरेजी, अंग्रेजी, अंग्रेजी

There are numerous words which are phonetically equivalent but vary in writing.

The word *school* in hindi can be written in different ways (स्कूल, सकूल, स्कुल) When information is searched for a single standard keyword school स्कूल and non standard Hindi phonetic equivalent keyword स्कुल 6.9 million results are shown by Google for former and 1.4 million for later. Hindi Language is influenced by the other regional languages which results in phonetic variety of words for example the English word school (स्कूल in Hindi) is pronounced and written as ISKOOL इस्कूल by the majority of population of India in different states. For the Hindi word ISKOOL इस्कूल more than two thousand results are found. Search engines should be capable of retrieving the results against phonetically equivalent words of keywords entered to search. User may use any keyword for searching and search engines should be capable to support all phonetically equivalent words. Following are randomly selected queries from the set of 50 queries tested on Google search engine

Following table 6.2.1 and graph 6.2.2 below show the results and precision offered by Google.

(Space for Table and graph 6.2.1 6.2.2)

In the above table it can be clearly seen that search engines return a handful of documents on various Hindi phonetically equivalent queries. It is observed that no particular standard exists for writing the keyword to fetch Hindi web data. For every phonetically equivalent keyword/s in the query variation in the results exist. I.e. a different set of documents are retrieved with least repetition. From the precision chart it is clearly observed that the degree of relevance for queries containing phonetically equivalent keywords is almost same or nearly equal. The native Hindi user may not be aware of the Phonetic issues in Hindi IR and may miss the relevant information of his/her use.

6.3 Words Synonyms

A word can express a myriad of implications, connotations, and attitudes in addition to its basic "dictionary" meaning. And a word often has near

synonyms that differ from it solely in these nuances of meaning. Choosing the right word can be difficult for people, as well as for the information retrieval system. For example the word (आभूषण) in Hindi (Ornament) in English, has following commonly spoken synonyms गहने, जेवर, अलंकार.

Table 6.3.1 and graph 6.3.2 has been presented below which shows the comparison of precision values against three search engines.

(Space for table 6.3.1 and graph 6.3.2)

From the experiments researchers observed that using Hindi keywords with their synonyms improves the information retrieval against a query in Hindi language.

Not only quantity of documents returned is affected but quality is also affected by using synonyms of Hindi keywords.

From the above table and graph it is be observed that documents returned by Google are more in quantity than other two search engines and least number of documents are returned by Guruji search engine the reason behind may be availability of less documents or poor indexing . However we are interested in quality of results than quantity, As far as quality of results is concerned it can be clearly seen that Google and Bing provide quality data than Guruji. And in the average case Google still stands first in the row that means precision values by Google are more than that of Bing and Guruji in this case. Thus it becomes clear that by changing a keyword into its synonym equivalent, results can be obtained. Therefore it is evident that synonyms of keywords play an important role in the process of Hindi information retrieval system.

6.4 Ambiguous words

Ambiguous words deflate the relevancy of the results. The examples mentioned below shows this aspect very clearly. Consider the following query.

(In English) →(Women like gold).

(In Hindi)→ (नारी को सोना पसंद है).

In this query the word सोना (Gold) is ambiguous as it has another meaning i.e. to sleep. In the context

of above query the word सोना is gold. But it can be also interpreted as women like to sleep.

Another Query: (In English) →(The common people's choice).

(In Hindi) →(आम लोगों की पसंद).

Here the word आम is ambiguous. The word आम in above query means common. However, In Hindi it also means mango. So the above query can be interpreted as "mango is people's choice". Various experiments have been done on this issue on three search engines to check their performance on handling of ambiguous word in a particular context.

We experimented on a sample set of 50 ambiguous queries and below we present five randomly selected ambiguous queries. In table (6.4.1) second column contains five queries in Hindi, third column holds the ambiguous keyword in one context and fifth column holds the same ambiguous keyword in other context. Fourth and sixth columns hold the meaning of queries in English with respect to the ambiguous keyword in context.

(Space for Table (6.4.1))

Ambiguous queries mentioned above in the table are tested for results against three search engines Google, Bing and Guruji. Results are shown below in tables 6.4.1, 6.4.2 and 6.4.3 as.

(Space for Tables 6.4.1, 6.4.2 and 6.4.3)

From the above results obtained in tables it is observed that all three search engines return documents without differentiating between the contexts of keyword in the query. In the above table the last column labeled as "other Context" holds the number of results which are not relevant to the query supplied or those documents which contains the keywords in other non required context. From the results it is clear that all search engines return documents in different contexts. Therefore it can be concluded that search engines underperform when supplied with ambiguous queries. Numbers in column labeled as "other Context" signifies the deviation from relevance. For example for query युद्ध में कुल विनाश (aggregate destruction in wars) the column "Other Context" for Google contains 5 documents for Bing contains 8 documents and for Guruji contains all 10 documents.

In another query सपेरो का फन (art of snake charmers) another context (Snake charmer's snake head) retrieved documents are expected to be in context (art) but from the above results obtained it can be seen that google returns all 10 documents in non required context (snake head) and Bing returns 9 documents where as Guruji fails to retrieve even a single document.

In the above scenario it becomes important for the search engines to address to the issue of ambiguity in keywords to obtain better results.

7 Discussion

We tested the performance of three search engines Google, Bing and Guruji for the challenges mentioned in the section (5). After the comparative analysis it is concluded that for *morphology*, query when supplied with root word yields maximum results. It is also evident that Google indexes the keyword in its root form and also lists the documents consisting of the morphological variants of the keywords. The other two engines Bing and Guruji do not list any morphological variant and hence they entail to have stemmer.

It is apparent from the table (6.2.1) that *phonetic* variation in Hindi keyword has a great impact on the performance of search engines. For each phonetically different word different set of results are obtained. Precision graph 6.2.2 also shows that the relevancy factor for all phonetic equivalent keywords contained in the queries is nearly equal for average case.

Word *synonyms* also play a major role in Hindi information retrieval process as it has been shown in table 6.3.1 above that a word with its synonyms when supplied to the search engines results in the retrieval of hand full of documents and none of the search engines is capable of listing a synonym of a keyword in the documents retrieved. However, the precision values of Google are better than other two search engines.

Ambiguous words in a query bring down the performance of search engines. None of the search engine is capable enough to handle the problem of ambiguity in query. It was observed that search results were far away from relevance and results obtained are out of context in almost all the cases.

8. Conclusions

In this paper we discussed the issues and problems which a user may face while finding Hindi information on web. We tested the parameters that affect the Hindi search on web. Search engines may have the performance and throughput problems if these parameters are implemented at root level. However this problem can be solved at interface level. Therefore in this direction we have developed software with a large scale Hindi database which is an interface between Hindi user and search engine. The software takes care of the Hindi phonetic variants, word synonyms and regional/foreign words that influence the Hindi Language. Complete description and implementation details will be reported shortly.

9 References

- [1]. Praveen Kumar, Shrikant Kashyap, Ankush Mittal Indian Institute of Technology, Roorkee, India. "A Query Answering System for E-Learning Hindi Documents" SOUTH ASIAN LANGUAGE REVIEW VOL.XIII, Nos 1&2, January-June,2003
- [2] S.K. Dwivedi and Parul Rastogi Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow, Uttar Pradesh, India "An Entropy Based Method for Removing Web Query Ambiguity in Hindi Language" Journal of Computer Science 4 (9): 762-767, 2008 ISSN 1549-3636 © 2008 Science Publications
- [3] S.K. Dwivedi and Parul Rastogi Rajesh kr. Gautam "Impact of language morphologies on search engine performance for hindi and English language" (IJCSIS) International Journal of Computer Science and Information Security, Vol. 8, No. 3, June 2010.
- [4] Prasad Pingali Jagadeesh Jagarlamudi Vasudeva Varma Language Technologies Research Centre IIIT, Hyderabad India. WebKhoj: Indian language IR from Multiple Character Encodings. URL: - www2006.org/programme/files/pdf/5503.pdf
- [5] Dimple Patel 1 and Devika P. Madalli 2 1 Mahatma Gandhi National Institute of Research and Social Action, Hyderabad, India 2 Documentation Research & Training Centre, Indian Statistical Institute, Bangalore, India "Information Retrieval in Indian Languages: A Case Study of Plural Resolution in Telugu Language". URL: - drtc.isibang.ac.in:8080/bitstream/handle/1849/396/054_p45_dimple.pdf
- [6] Government of India Ministry of Home Affairs Department of Official Language Technical Cell "Office Memorandum No 12015/7/2008-OL (TC.)
- [7] Priyendra S. Deshwal Kalyanmoy Deb Department of Computer Science and Engineering Indian Institute of Technology, Kanpur. Kanpur PIN-208016, India. "Design of an Optimal Hindi Keyboard for Convenient and Efficient Use" URL: - www.iitk.ac.in/kangal/papers/k2003004.pdf

[8] Ranjan Srivastava, Chief of Bureau, Prabhat Khabar, (Friday, April 28, 2006) "The future of Hindi on the Internet" <http://www.raftaar.in/thehoot.htm>.

[9] Rajeev Rathor Master of Engineering Thesis Thapar University." Patiala Morphological POS Tagger for Hindi Language" URL: -
<http://dspace.thapar.edu:8080/dspace/bitstream/10266/554/1/Rajeev+Thesis+report.pdf>

[10] GANAPATHIRAJU Madhavi, BALAKRISHNAN Mini, BALAKRISHNAN N. REDDY Raj "Om: One tool for many (Indian) languages" Journal of Zhejiang University SCIENCE ISSN 1009-3095.

[11] Vishal Goyal, Ph.D. Development of a Hindi to Punjabi Machine Translation System
A Doctoral Dissertation Volume 10: 10 October 2010 ISSN 1930-2940 Language in India www.languageinindia.com

S. No	Query in Hindi	Meaning In English	S.NO	Query in Hindi	Meaning In English
1	भारत वर्षावन	Indian rain forest	6.2	पक्षियों की प्रजातियां	Bird's species
1.1	भारतीय वर्षा वनों	Indian rain forests	7	कृषि समस्या	Agriculture problem
2	हवाई दुर्घटना का कारण	Reason for air crash	7.1	कृषि समस्याओं	Agriculture problems
2.1	हवाई दुर्घटनाओं का कारण	Reason for air crashes	8	कीटनाशक के इस्तेमाल	Use of pesticide
3	भारत में बोली जाने वाली भाषा	Language spoken in India	8.1	कीटनाशकों के इस्तेमाल	Use of pesticides
3.1	भारत में बोली जाने वाली भाषाएँ	Languages spoken in India	9	मानसिक रोग	Mental illness
3.2	भारत में बोली जाने वाली भाषाओं	Languages spoken in India	9.1	मानसिक रोगियों	Mental patients
4	विलुप्त होने पर झील	Lake on the verge of extinction	9.2	मानसिक रोगी	Mental Patient
4.1	विलुप्त होने पर झीलें	Lakes on the verge of extinction	10	ग्रामीण विकास योजना	Policy for village
5	प्राकृतिक आपदा	Natural calamity	10.1	ग्रामीण विकास योजनाओं	Policies for village
5.1	प्राकृतिक आपदाएँ	Natural calamities	10.2	ग्रामीण विकास योजनाएं	Policies for village
6	पक्षी की प्रजाति	Bird species	11	प्रमुख कृषि केंद्र	Major agricultural office
6.1	पक्षियों की प्रजाति	Bird's species	11.1	प्रमुख कृषि केन्द्रों	Major agricultural offices

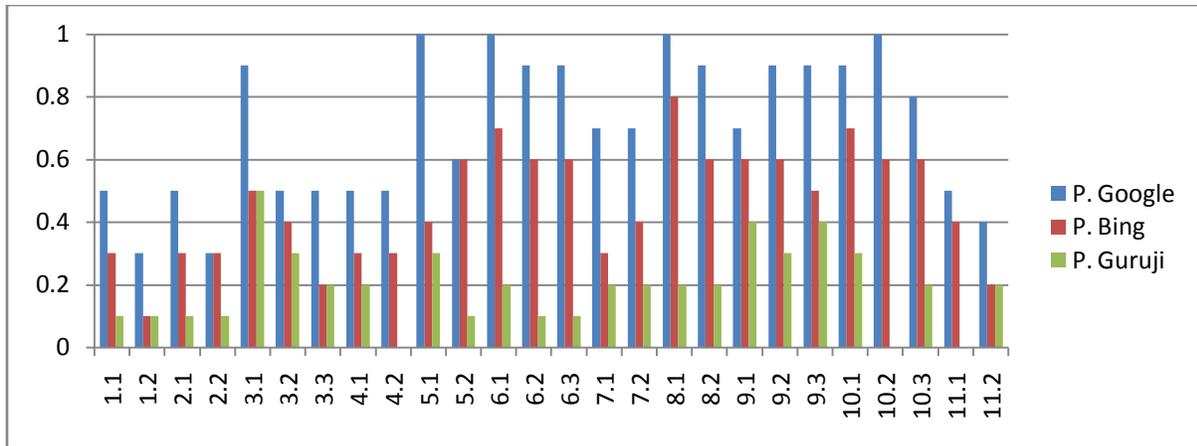
Table 6.1 List of Hindi queries

S. No	Root word/s	Listing of Keywords			Morphological variants	Documents Returned			
		Google	Bing	Guruji		Google	Bing	Guruji	
1	भारत , वर्षा वन	भारत, वर्षावन, वर्षा वन, वन	वर्षा वन, वन	भारत ,वर्षा वन, वन	1.1	भारत , वर्षा वन	50,500	4,680	485
					1.2	भारतीय वर्षा वनों	40,400	680	61
2	दुर्घटना	दुर्घटना,दुर्घटनाओं,	दुर्घटना	दुर्घटना	2.1	दुर्घटना	133,000	2,410	284
					2.2	दुर्घटनाओं	117,000	420	23
3	भाषा	भाषा, भाषाओं,	भाषा	भाषा	3.1	भाषा	161,000	8,330	961
					3.2	भाषाएँ	6,200	935	188
					3.3	भाषाओं	6,090	441	356
4	झील	झील, झीलें,	झील	झील	4.1	झील	4,740	278	25
					4.2	झीलें	1,270	28	1
5	आपदा	आपदा, आपदाओं,	आपदा	आपदा	5.1	आपदा	102,000	4,030	410
					5.2	आपदाएँ	1,160	64	20
6	पक्षी,प्रजाति	पक्षी, पक्षियों, प्रजाति, प्रजातियों, प्रजातियां	पक्षी, प्रजाति	पक्षी, प्रजाति	6.1	पक्षी ,प्रजाति	48,200	1,670	98
					6.2	पक्षियों, प्रजाति	47,600	1,150	84
					6.3	पक्षियों ,प्रजातियां	33,800	747	25
7	समस्या	समस्याएं,समस्याओं, समस्या	समस्या	समस्या	7.1	समस्या	584,000	30,200	1,889
					7.2	समस्याओं	584,000	7,150	1,356
8	कीटनाशक	कीटनाशक, कीटनाशकों	कीटनाशक	कीटनाशक	8.1	कीटनाशक	36,300	1,360	333
					8.2	कीटनाशकों	35,800	800	270
9	रोग	रोगों, रोग	रोग	रोग	9.1	रोग	205,000	21,600	1,423
					9.2	रोगियों	128,000	3,280	239
					9.3	रोगी	112,000	6,280	647
10	योजना	योजनाओं,योजना	योजना	योजना	10.1	योजना	673,000	18,500	3,343
					10.2	योजनाओं	669,000	6,020	990
					10.3	योजनाएं	673,000	2,860	416
11	केंद्र	केंद्रीय, केंद्र,	केंद्र,	केंद्र,	11.1	केंद्र	261,000	11,300	655
					11.2	केंद्रों	29,500	1,850	105

Table 6.1.1 Effect of morphological factors on Hindi queries

S. No	Query	Precision@ 10			S.NO	Query	Precision@ 10		
		Google	Bing	Guruj i			Google	Bing	Guruji
1.1	भारत वर्षा वन	0.5	0.3	0.1	6.3	पक्षियों की प्रजातियां	0.9	0.6	0.1
1.2	भारतीय वर्षा वनों	0.3	0.1	0.1	7.1	कृषि समस्या	0.7	0.3	0.2
2.2	हवाई दुर्घटना का कारण	0.5	0.3	0.1	7.2	कृषि समस्याओं	0.7	0.4	0.2
2.2	हवाई दुर्घटनाओं का कारण	0.3	0.3	0.1	8.1	कीटनाशक के इस्तेमाल	1	0.8	0.2
3.1	भारत में बोली जाने वाली भाषा	0.9	0.5	0.5	8.2	कीटनाशकों के इस्तेमाल	0.9	0.6	0.2
3.2	भारत में बोली जाने वाली भाषाएँ	0.5	0.4	0.3	9.1	मानसिक रोग	0.7	0.6	0.4
3.3	भारत में बोली जाने वाली भाषाओं	0.5	0.2	0.2	9.2	मानसिक रोगियों	0.9	0.6	0.3
4.1	विलुप्त होने पर झील	0.5	0.3	0.2	9.3	मानसिक रोगी	0.9	0.5	0.4
4.2	विलुप्त होने पर झीलें	0.5	0.3	0	10.1	ग्रामीण विकास योजना	0.9	0.7	0.3
5.1	प्राकृतिक आपदा	1	0.4	0.3	10.2	ग्रामीण विकास योजनाओं	1	0.6	0
5.2	प्राकृतिक आपदाएँ	0.6	0.6	0.1	10.3	ग्रामीण विकास योजनाएं	0.8	0.6	0.2
6.1	पक्षी की प्रजाति	1	0.7	0.2	11.1	प्रमुख कृषि केंद्र	0.5	0.4	0
6.2	पक्षियों की प्रजाति	0.9	0.6	0.1	11.2	प्रमुख कृषि केन्द्रों	0.4	0.2	0.2

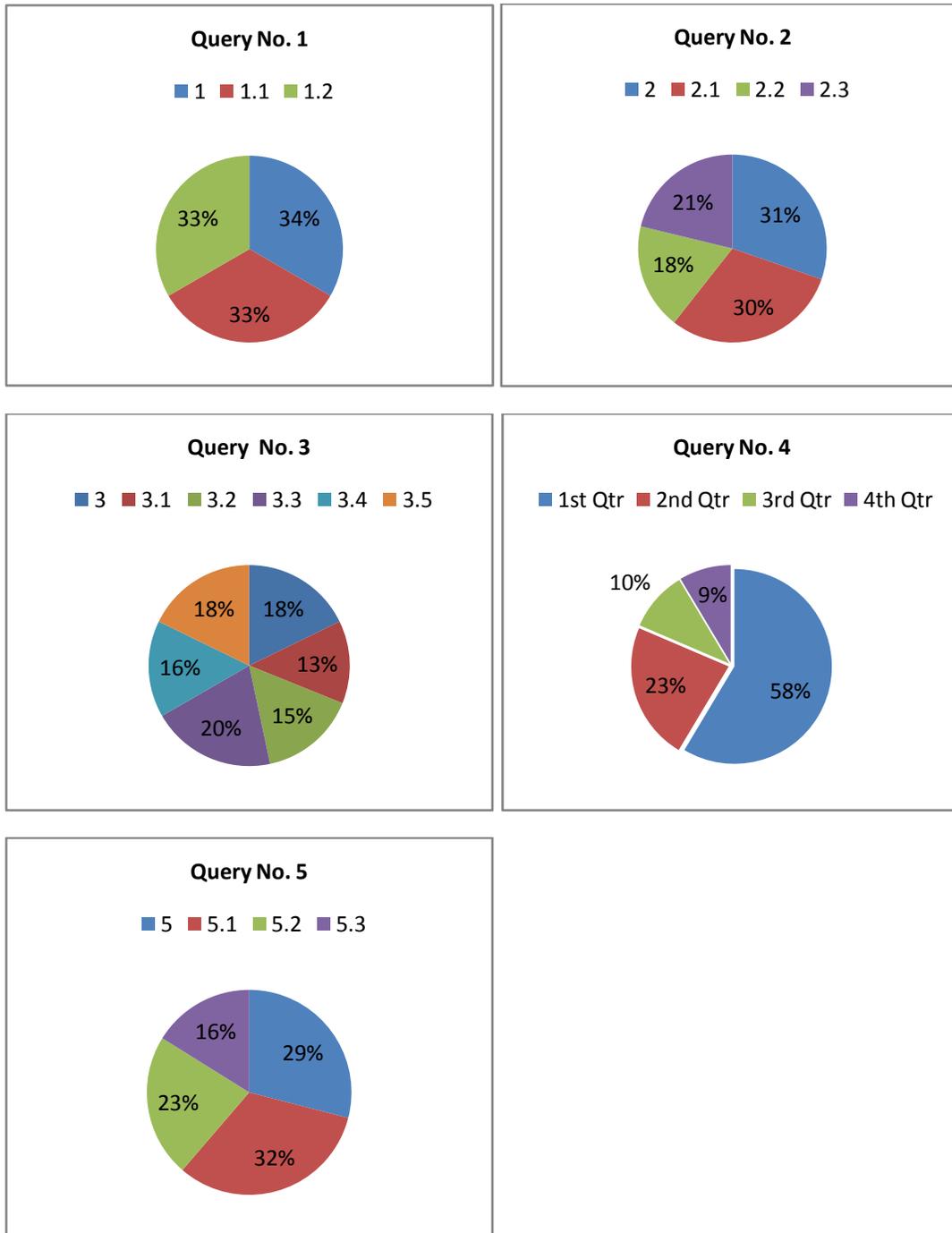
Table 6.1.2 precision values of the three search engines



Graph 6.1.3 precision values of the three search engines

Hindi Query With Bold Standard Keywords	Phonetic variations of the Keywords		Google Results for query having keywords	No. of Results	Precision @10	
सब्जियों में जहरीले पदार्थ	सब्जियों सब्जियों,	जहरीले, जहरिले	सब्जियों, जहरीले	97	0.9	
			सब्जियों, जहरीले	632	0.9	
			सब्जियों, जहरीले	194	0.9	
आसमान छूती महंगाई	आसमां,	महंगाई , मेहंगाई	आसमान महंगाई	35300	1.0	
			आसमान महंगाई	1040	1.0	
			आसमां महंगाई	14	0.6	
			आसमां महंगाई	563	0.7	
भ्रष्टाचार से आजादी	भ्रष्टाचार भ्रष्टाचार	आजादी	भ्रष्टाचार आजादी	211,000	0.8	
			भ्रष्टाचार आजादी	214	0.6	
			भ्रष्टाचार आजादी	447	0.7	
			भ्रष्टाचार आजादी	1,090,000	0.9	
			भ्रष्टाचार आजादी	1,040	0.7	
			भ्रष्टाचार आजादी	1,190	0.8	
अन्ना हजारे का आन्दोलन	अनना	हजारे	आंदोलन आंदोलन	अन्ना हजारे	84,700	0.3
				अन्ना हजारे आंदोलन	85,100	0.8
				अन्ना हजारे का आंदोलन	78	0.6
				अनना हजारे का आन्दोलन	399	0.5
				अन्ना हजारे का आन्दोलन	3,260,000	1.0
बेरोजगारी समस्या समाधान	बेरोजगारी बरोजगारी बेरोजगारी		बेरोजगारी	9,650	0.9	
			बेरोजगारी	80,600	1.0	
			बरोजगारी	170	0.7	
			बेरोजगारी	30	0.5	

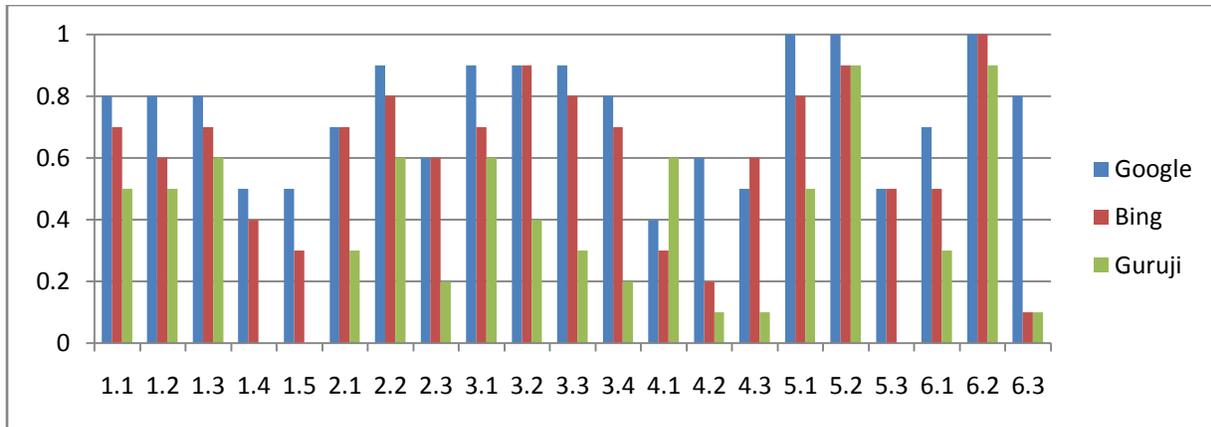
Table 6.2.1 results of three search engines on Phonetic nature of Hindi language



Graphs 6.2.2 Precision Charts for Phonetic nature of Hindi language

S. NO	Query	Standard Hindi Word/s	Synonyms		Documents Returned					
					Google	Per. @10	Bing	Per. @10	Guruj i	Per. @10
1	सोने के आभूषण	आभूषण/ गहने	1.1	सोने के आभूषण	217,000	0.8	3,250	0.7	381	0.5
			1.2	सोने के गहने	188,000	0.8	2,590	0.6	389	0.5
			1.3	सोने के जेवर	78,900	0.8	1,670	0.7	311	0.6
			1.4	सोने के अलंकार	9,490	0.5	633	0.4	70	0
			1.5	सोने के आभरण	493	0.5	38	0.3	1	0
2	काले बादल	बादल	2.1	काले बादल	233,000	0.7	7,510	0.7	733	0.3
			2.2	काले मेघ	40,700	0.9	1,500	0.8	99	0.6
			2.3	काले जलधर	1,570	0.6	54	0.6	2	0.2
3	स्त्री सशक्तिकरण	स्त्री, नारी	3.1	स्त्री सशक्तिकरण	9,950	0.9	1,570	0.7	760	0.6
			3.2	नारी सशक्तिकरण	29,300	0.9	1,910	0.9	736	0.4
			3.3	महिला सशक्तिकरण	96,300	0.9	5,160	0.8	1,091	0.3
			3.4	औरत सशक्तिकरण	7,670	0.8	680	0.7	510	0.2
4	सिकंदर का अंहकार	अंहकार	4.1	सिकंदर का अंहकार	1,990	0.4	18	0.3	60	0.6
			4.2	सिकंदर का अभिमान	2,400	0.6	304	0.2	16	0.1
			4.3	सिकंदर का घमंड	495	0.5	54	0.6	9	0.1
5	वृक्ष लगाओ	वृक्ष	5.1	वृक्ष लगाओ	6,960	1	698	0.8	29	0.5
			5.2	पेड़ लगाओ	13,400	1	1,080	0.9	143	0.9
			5.3	दरख्त लगाओ	481	0.5	19	0.5	0	0
6	आँख दान	आँख	6.1	आँख दान	34,000	0.7	3,690	0.5	312	0.3
			6.2	नेत्र दान	77,500	1	3,240	1	159	0.9
			6.3	चक्षु दान	2,450	0.8	427	0.1	36	0.1

Table 6.3.1 effect of word synonyms on Hindi IR



Graph 6.3.2 comparison of precision values against three search engines

S.No	Query	For keyword as	In English	For keyword as	In English
1	नारी को सोना पसंद है	सोना (Gold)	Women like Gold	सोना (To sleep)	Women like to sleep
2	आम लोगों की पसंद	आम (common)	Common man's choice	आम (Mango)	Mango is people's choice
3	बाल विकास और पोषण	बाल (Children)	Child Development and Nutrition	बाल(Hair)	Hair Development and Nutrition
4	सपेरोँ का फन	फन (Art)	Art of snake charmers	फन(Snake head)	Snake charmer's snake head
5	युद्ध में कुल विनाश	कुल (Aggregate)	Aggregate destruction in wars	कुल(family)	Destruction of families in war

Table 6.4.1 List of randomly selected ambiguous queries

Query	Ambiguous keyword	Documents returned	Google				Other Context
			Results Found				
			Context		Context		
1	सोना	50,800	Gold	5	To sleep	2	3
2	आम	488,000	Common	3	Mango	3	4
3	बाल	2,900,000	Children	7	Hair	3	0
4	फन	184	Art	0	Snake head	10	0
5	कुल	17,800	Aggregate	2	Family	3	5

Table 6.4.2 Ambiguity test for Google

Query	Ambiguous keyword	Documents returned	Bing				Other Context
			Results Found				
			Context		Context		
1	सोना	2,680	Gold	2	To sleep	2	6
2	आम	17,800	Common	3	Mango	3	4
3	बाल	4,030	Children	6	Hair	2	2
4	फन	25	Art	0	Snake head	9	1
5	कुल	1,900	Aggregate	0	Family	2	8

Table 6.4.3 Ambiguity test for Bing

Query	Ambiguous keyword	Documents Returned	Guruji				Other Context
			Results Found				
			Context		Context		
1	सोना	109	Gold	0	To sleep	0	10
2	आम	6,756	Common	3	Mango	0	7
3	बाल	635	Children	5	Hair	2	3
4	फन	No Results Found	Art	n/a	Snake head	n/a	n/a
5	कुल	84	Aggregate	0	Family	0	10

Table 6.4.4 Ambiguity test for Guruji

Gain Flattening and Noise Figure Analysis of a Dual Stage Bowtie WDM EDFA Configuration in C-Band

Ricky Anthony, Sambhunath Biswas

Abstract— The paper presents an improved gain flattening and noise figure analysis in the C-band regime using an equalization filter of a dual stage single mode polarization maintaining bowtie erbium doped fiber amplifier. The configuration uses two in-line 980nm laser pumps. The gain and noise figure variation with fiber length, pump power and temperature of the system for C-band communication have been investigated. With equalization filter, a gain and noise figure flatness (p-p) of just 1.1dB and 2.3dB respectively, in 1530-1565 nm BW, for optimized 150 mW and 250 mW pump lasers was obtained. The gain has also shown temperature dependent variation, with minimum fluctuation of 0.71dB at 20°C. Based on results, the system feasibility as in C-band communication has been discussed.

Index Terms— Erbium doped fiber amplifier, Bowtie polarization maintaining fiber, Giles model, McCumber's equations, C-band, Gain flattening, Noise figure, Equalization filter.

1 INTRODUCTION

The ever increasing demand for high speed data communication over long distances with minimal of losses have paved way for the erbium doped fiber amplifiers (EDFA) as in-line and pre-optical amplifiers and data carrying medium, hence becoming an obvious replacement for electronic regenerative repeaters for low amplified spontaneous noise (ASE) level. It has also shown wide bandwidths, lower noise figure and polarization insensitivity. The EDFA however has been successfully commercialized as a potential mean for data communication in conventional or C-band (1530-1565nm), long or L-band (1570-1620nm) region, and recently possibilities of short or S-band (1460-1530nm) as an effective communication channel have been explored [1, 2].

The introduction of wavelength division multiplexing (WDM) with rare earth doped amplifier such as EDFAs as power boosters and pre-amplifiers have increased the channel capacity to as high as 69.1 Tb/s over 240 km [3]. Figure 1 and 2 shows the standard co-directional EDFA pumped with 980nm laser and the EDFA gain spectrum obtained using EDFA GainMaster™ simulation tool, respectively. Such a wavelength selective variation of gain in case of a multiple wavelength input signal with same power would lead to large output power differences, giving poor signal to noise ratio (SNR) among various channels. This phenomenon gets particularly prominent during small signal analysis. Hence, this largely restricts EDFA application in a narrow band region which is undesirable for

of equalization filters, fiber Bragg gratings (FBG) or Mach-Zehnder interferometer in the system [4]. Other approaches include inherent gain flattened EDFA such as asymmetric twin core (ATC) EDF which has a gain spectrum flattened within ± 1 dB over 32 nm range [5].

In case of cascaded WDM systems, use of narrow band filters after each stage would increase ASE and saturate the subsequent amplifiers and hence reduce the gain in each stage. The noise associated with EDFA is due to ASE (forward and backward), which has a spectrum almost the same as the gain spectrum and needs to be reduced, especially the forward ASE, which propagates in the direction of the signal, to make EDFAs more incompatible for long haul communication. Figure 3 shows the noise figure (NF) associated with co-directional EDFA configuration as a function of wavelength.

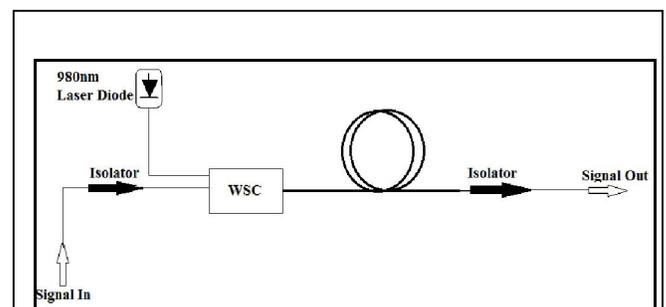


Fig.1. A standard co-directional 980nm pumped EDFA configuration.

- Ricky Anthony is currently pursuing Masters Degree program in Electronics and Communication Engineering in Heritage Institute of Technology, India, PH-09748998348. E-mail: ricky.j.anthony@gmail.com.
- Dr.S.N.Biswas is the Deputy Director and Senior Professor of Heritage Institute of Technology, India, PH-09231509082. E-mail: s.biswas@rediff.com.

optical communication. It can be compensated with inclusion

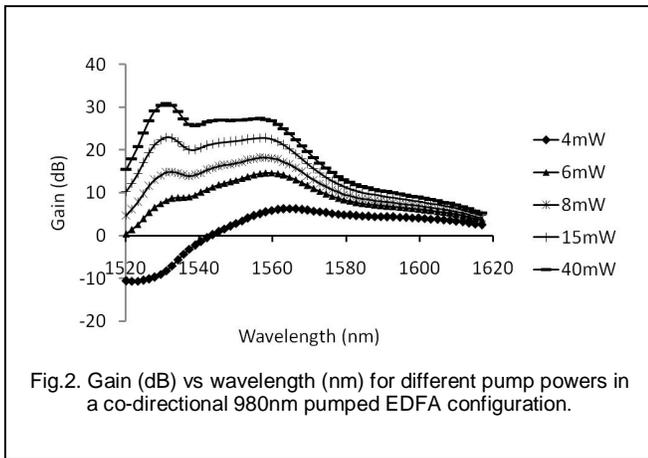


Fig.2. Gain (dB) vs wavelength (nm) for different pump powers in a co-directional 980nm pumped EDFA configuration.

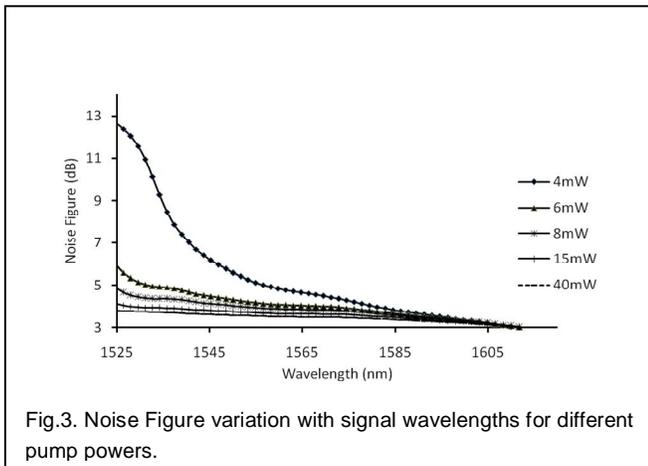


Fig.3. Noise Figure variation with signal wavelengths for different pump powers.

2 EDFA CHARACTERISTICS

Reduction of fiber losses and gain variation depends on the erbium doped fiber and its characteristics considerably. The dual stage configuration presented in this paper is based on specialty polarization maintaining (PM) bowtie single mode (SM) erbium fiber, shown in figure 4. Even though erbium doped fiber (EDF) are polarization independent, but some of the Er^{+3} dopant ions may have certain polarization, giving a polarization dependent gain (PDG). This is because there is a polarization offset between the input signal laser and the pump laser. Other such PM fibers include PANDA and elliptical jacket fiber.

First developed by Optical Fiber Group at University of Southampton in 1982 [6], the bowtie based PM induces birefringence when the fiber core experiences tension due to fiber drawing. This is because the core is guarded by boron-doped glass bow-ties which can shrink more compared to its silica based surrounding cladding. The birefringence produced, allows the incident light along the "slow axis" to travel at lower velocity compared to the light incident along the "fast axis". This velocity change along the two axes makes the cross-coupling of light difficult and rare, maintaining the polarization. Hence, greater is the stress applied; larger is the difference in the propagation constant. Both bowtie and PANDA today find extensive use in telecommunication industry as optical modulators, sensors, interferometers and gyroscopes.

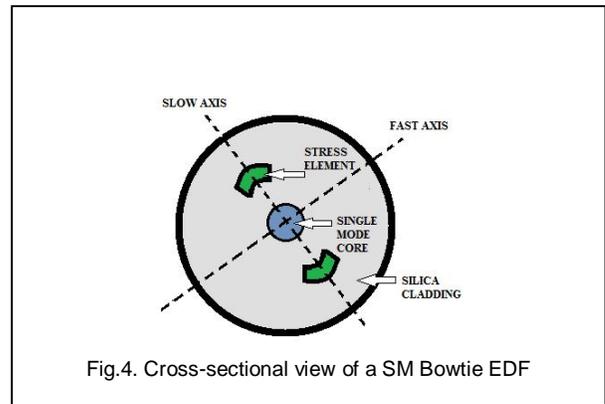


Fig.4. Cross-sectional view of a SM Bowtie EDF

The following table gives all the specifications associated with the bowtie EDF used for the analysis of dual stage gain flattened configuration.

TABLE 1.

BOWTIE EDFA SPECIFICATION

Parameters	Specifications*
Cut-off wavelength	860-960 nm
Numerical Aperature	0.22-0.26
Mode Field Diameter	3.5µm at 980nm 6.0µm at 1550 nm
Pump Absorption at 980 nm	10dB/m (minimum)
Signal Absorption at 1530 nm	12-27dB/m
Attenuation at 1200 nm	Less than 20 dB/m
Fiber Diameter	125µm approx.
Coating Diameter	245µm approx.
Coating Type	Dual Acrylate

Obtained from Fibercore DHB 1500 datasheet.* nm = nanometer, µm = micrometer, dB = decibel, m = meter.

3 EDFA MODELING FOR TWO LEVEL SYSTEM

Modeling of EDFA for simulation and performance analysis is largely based on the landmark work by C.R.Giles and E.Desurvire [7] in 1992. The EDFA model utilizes giles parameters which comprise of the wavelength dependent, absorption coefficient $\alpha(\lambda)$ and gain coefficient $g^*(\lambda)$. The absorption coefficient is also dependent on the erbium ion concentration, and calculated with all laser active ions are in ground state. The gain coefficient or the emission spectra on the others hand is calculated with laser active ions in the excited level. Both $\alpha(\lambda)$ and $g^*(\lambda)$ are dependent on absorption cross section, $\sigma_a(\lambda)$ and emission cross-section, $\sigma_e(\lambda)$ respectively and given by the equation:

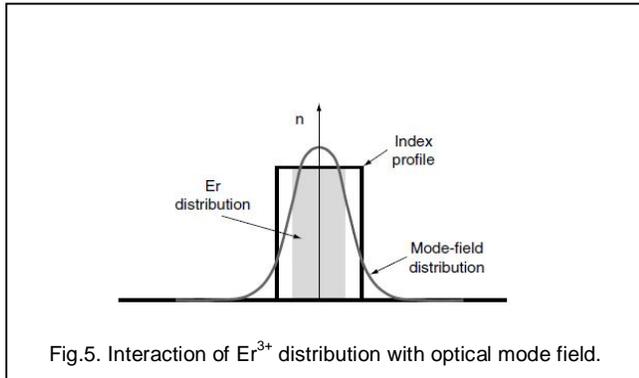
$$\alpha_k(\lambda_k) = \sigma_a(\lambda_k) \cdot \int_0^{2\pi} \int_0^\infty i_k(r, \phi) \cdot n_t(r, \phi, z) \cdot r dr d\phi \quad (1)$$

$$g_k(\lambda_k) = \sigma_e(\lambda_k) \cdot \int_0^{2\pi} \int_0^\infty i_k(r, \phi) \cdot n_t(r, \phi, z) \cdot r dr d\phi \quad (2)$$

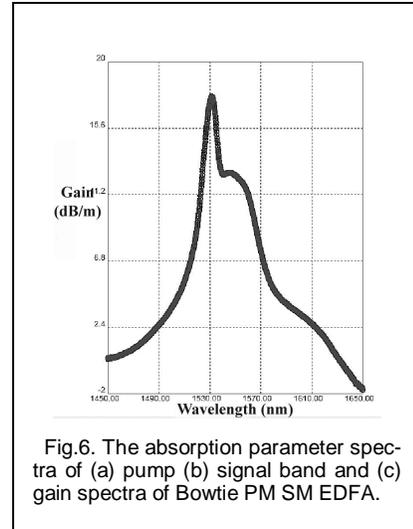
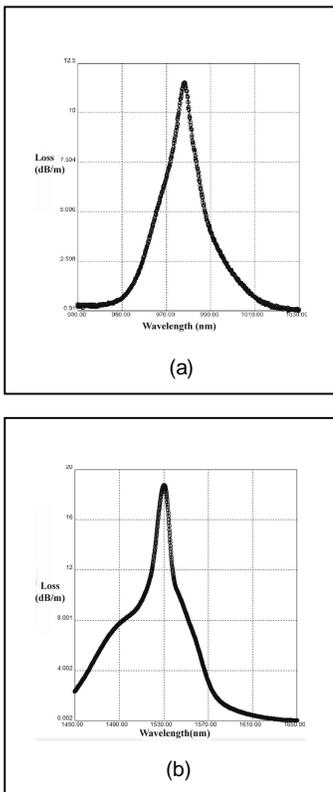
$$\alpha_k(\lambda_k) = \Gamma(\lambda_k) \cdot \bar{n}_t \cdot \sigma_a(\lambda_k) \quad (3)$$

$$g_k(\lambda_k) = \Gamma(\lambda_k) \cdot \bar{n}_t \cdot \sigma_e(\lambda_k) \quad (4)$$

Where $\Gamma(\lambda)$ is the overlap factor between the optical mode field and erbium ions (shown in fig.5), which results in stimulation absorption or emission from Er^{3+} transitions and n_t is the total Er^{3+} ion density, which can be measured experimentally using X-ray photoelectron spectroscopy (XPS), also known as electron spectroscopy for chemical analysis (ESCA) [8]. The overlap factor is a function of with \square and r in spherical coordinate system with \square as constant and 'b' is the radius of the Er^{3+} doped region.



These equations are however based on the assumption that parasitic losses such as scattering losses are minimal, Er^{3+} is uniformly distributed in the core and gain is dependent on laser transition only. Fig.6 (a), (b) and (c), shows the absorption parameter for both signal band and pump band, and gain parameter spectra of PM single mode (SM) bowtie EDF obtained from simulation. It is also assumed that no ion-ion interaction is present.



The Giles algorithm solves the propagation equation (5) by integrating in both backward and forward direction in an iterative manner until the solution converges.

$$\frac{dP_k(z)}{dz} = u_k \cdot P_k(z) \cdot \left(g_k(v_k) + \alpha_k(v_k) \cdot \frac{\bar{n}_2}{\bar{n}_1} - \alpha_k(v_k) - l_k \right) + u_k \cdot P_{ok} \cdot g_k(v_k) \cdot \frac{\bar{n}_2}{\bar{n}_1} \quad (5)$$

Where, k signifies a definite signal, and u_k is the beam propagation direction with values 1 and -1 for forward and backward propagation and P_{ok} gives the contribution of the spontaneous emission from the local excited state population n_2 and is equal to $mh\nu_k \Delta v_{k,}$ where m defines the number of polarization modes supported by the fiber.

The fiber saturation parameter ζ ($m^{-1}s^{-1}$) is measured by reducing the 980nm laser pump to the power where it reaches its saturation value. For an equivalent radius of doped region, b_{eff} , it is given by:

$$\zeta = \pi \cdot b_{eff}^2 \cdot n_t / \tau \quad (6)$$

Under steady state condition of the EDFA with uniform distribution for both upper and lower excited states, the steady state equation for population density at the upper level would be [7]:

$$\frac{\bar{n}_2}{\bar{n}_1}(z) = \frac{\sum_{k=1}^n \frac{P_k(z) \cdot \alpha_k v_k}{h \cdot v_k \cdot \zeta}}{1 + \sum_{k=1}^n \frac{P_k(z) \cdot (\alpha_k(v_k) + g_k(v_k))}{h v_k}} \quad (7)$$

The intrinsic fiber power saturation, $P_{sat}(\lambda)$ is given by:

$$P_{sat}(\lambda) = \frac{h\nu A}{[\sigma_a(\lambda) + \sigma_e(\lambda)]\Gamma(\lambda)\tau} \quad (8)$$

Where, h is the Planck constant, ν is the frequency for the signal or the laser pump, A is the area of the doped region and τ is the lifetime of the metastable level.

The major source of noise in case of EDFA is the amplified spontaneous emission (ASE). During EDFA band transition, the stimulated photons are coherent in nature which contributes to amplification; however a photon in the excited state which is not stimulated within a lifetime of 10ms of the excited state contributes to spontaneous emission, which is incoherent in nature and is responsible for ASE noise (forward and backward). For a cascaded EDFA configuration, the ASE keeps on increasing at each successive step, hence reducing the overall gain of the system. This can be attenuated with the introduction of optical filters. But in long haul systems, with 200 amplifiers [9], the 1530nm have shown large attenuation whereas 1560nm peak have shown sharp rise. This shift in ASE power is attributed to wavelength-dependence nature of absorption and emission cross-sections, and saturations of the amplifier at each stage. Recent surveys have shown utilization of backward ASE power for simultaneous gain flattening at C and L-bands [10, 11].

4 RESULTS

The EDFA configuration (as shown in fig.7) under analysis had a multiple source with frequencies ranging from 1530nm to 1565nm with 15 channels of 0.01mW/channel simulated using EDFA GainMaster™. It was coupled with 980 nm pump laser by means of wavelength division multiplexing (WDM) with 200GHz spacings according to ITU standards, to PM SM bowtie EDFA. The output was again coupled to another 980nm pump laser using WDM and its output is filtered using an equalization filter to obtain a relatively flattened gain and NF. Optical isolators were used to avoid backward ASE power back into the source.

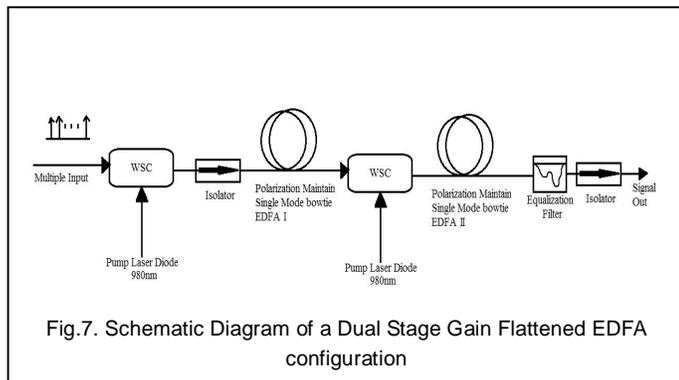


Fig.7. Schematic Diagram of a Dual Stage Gain Flattened EDFA configuration

4.1 Gain Flattened Curve

The gain spectrum for various combinations of pump powers for both the pump lasers was simulated and optimized (shown in Fig.13). It was found that with 150 mW, 250 mW pump power lasers I and II respectively with 5 m EDFA lengths, gave a flat gain spectrum variation of just 1.1dB for the entire C-band. As the pump power decreased, the gain spectrum showed larger variation in the C-band region. The table 2 gives the comparison of the gain variation for different pump powers. Such a flat gain system finds application in CATV and analog systems.

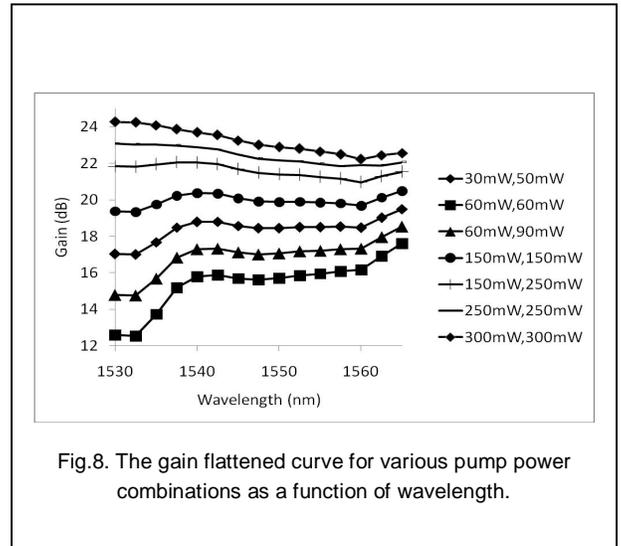


Fig.8. The gain flattened curve for various pump power combinations as a function of wavelength.

TABLE 2

GAIN CHARACTERISTICS FOR DIFFERENT PUMP POWERS

Sl. No	980nm Pump Power (I)	980nm Pump Power (II)	Gain Flatness (p-p) in dB
1.	30 mW	50 mW	2.94
2.	60 mW	60 mW	5.0
3.	60 mW	90 mW	3.78
5.	150 mW	150 mW	1.14
6.	150 mW	250 mW	1.10
7.	250 mW	250 mW	1.17
8.	300 mW	300 mW	2.05

4.2 Noise Figure Analysis

The primary source of noise in any doped fiber amplifier (DFA) system is because of the presence of amplified spontaneous emission (ASE), which extends the entire gain spectrum. It was assumed in the simulation that ASE extends from 1520nm-1620nm. With the inclusion of an equalizer filter, the NF was found to be as low as 7.48dB which increased to 9.80dB over the entire C-band for gain-flattened optimized pump powers of 150mW and 250mW each (fig 9). This also indicates that for C plus L-band configuration would require two individual filters, since the NF increases as the wavelength approaches L-band regime. Comparison Table 3 gives the NF variation for different pump powers. For a cascaded system, the ASE accumulates at each stage reducing the overall gain at the output.

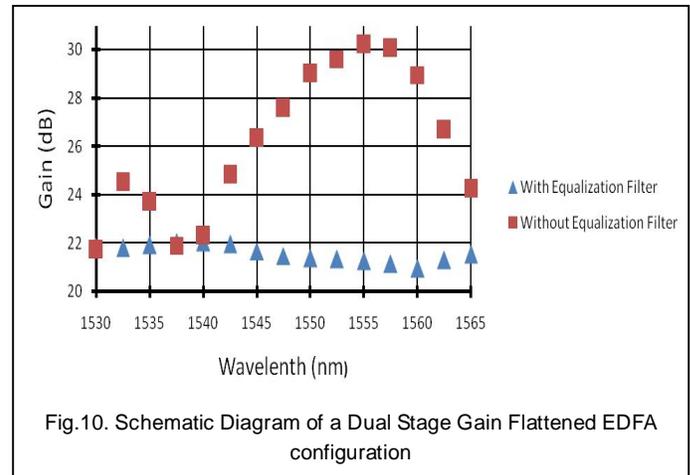
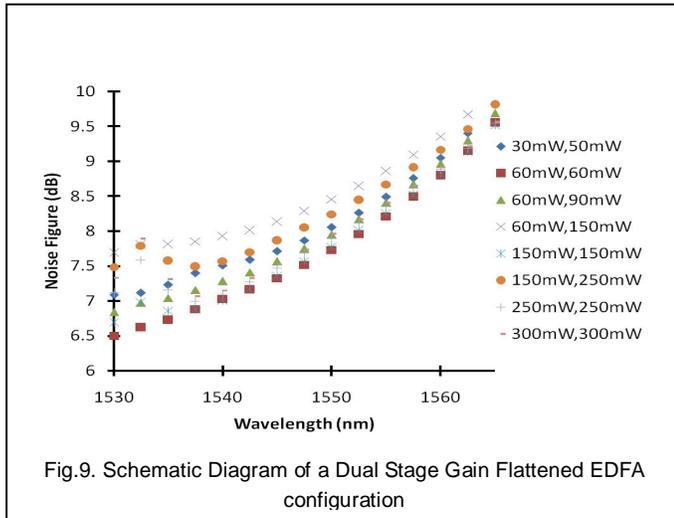


Fig.10. Schematic Diagram of a Dual Stage Gain Flattened EDFA configuration

TABLE 3

NOISE FIGURE CHARACTERISTICS FOR DIFFERENT PUMP POWERS

Sl. No.	980nm Pump Power (I)	980nm Pump Power (II)	NF Flatness (p-p) in dB
1.	30 mW	50 mW	2.73
2.	60 mW	60 mW	3.0
3.	60 mW	90 mW	2.86
5.	150 mW	150 mW	2.84
6.	150 mW	250 mW	2.30
7.	250 mW	250 mW	2.42

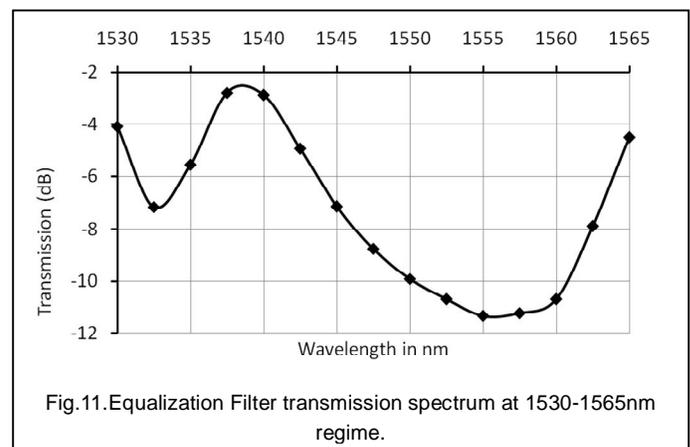


Fig.11. Equalization Filter transmission spectrum at 1530-1565nm regime.

4.3 Equalization Filter Characteristics

The gain spectrum can be controlled either by changing fiber host properties, improving measurement techniques and employing new EDFA system designs. This paper uses a dual stage design, gain flattened by an optical filter. An optical gain equalization filter such as fiber grating filter is one of the most common ways to reduce the gain peaks to provide an equalized output. In the dual stage configuration under consideration, the primary peak obtained in the absence of filter for a 150 mW, 250 mW pumped laser and 5m EDF length is 24dB and, the secondary peak occurs at 30dB (fig 10). These peaks are reduced considerably in the presence of filter as shown in figure obtained from simulations. The transmission spectrum of the filter is shown in fig 11, which clearly indicates that the filter has spectrum inverse to that of the two stage EDFA system. The NF depends on the position of the filter in the system, and so the equalization filter was strategically chosen to be placed at the end of the second stage.

4.4 Gain Dependence on EDF Length and Temperature

The EDFA gain is length dependent, for a given pump power, the gain reaches a maximum value before decreasing. This is because after propagating through a certain distance in the optical fiber, the pump power reduces due power absorption. Hence designing of an EDFA requires prior consideration of pump power and EDFA length. Fig 12 shows the gain dependence secondary EDFA for different channel wavelength. As seen from the graph, channels with higher frequency fall off quickly compared to channels with lower frequency. This becomes a limiting factor for WDM based optical communication.

The C-Band flattened gain undergoes temperature variation, according to the changes in ion energy levels in accordance to McCumbers equations [12]:

$$\frac{g^*(\lambda)}{\alpha(\lambda)} = \exp\left[\frac{hc}{kT}\left(\frac{1}{\lambda} - \frac{1}{\lambda_0}\right)\right] \quad (9)$$

Where, k is boltzman's constant, h is Planck's constant, T is temperature in degree kelvin and λ_0 is the cross-over wavelength. During the simulation, source and detector was considered to be temperature independent. Such temperature fluctuations have shown to have affected signal to noise ratio (SNR) and increased bit error rates (BER) for some channels

compared to other [13]. The fig 13 shows the gain fluctuation over a temperature range of 0°C to 50°C over the C-band. The temperature coefficient of an EDFA is positive below 1540nm and negative above it. And for conventional equalization filters the temperature coefficient is positive below 1535 nm and negative above it. For an equalization filter based gain flattened EDFA configuration, above 1535nm the gain spectra of EDF and filter transmission spectra cancel out, whereas below this wavelength they reinforce, giving overall gain spectra as obtained in figure 18 from simulation. Table 4 shows the relative gain variations at different temperatures.

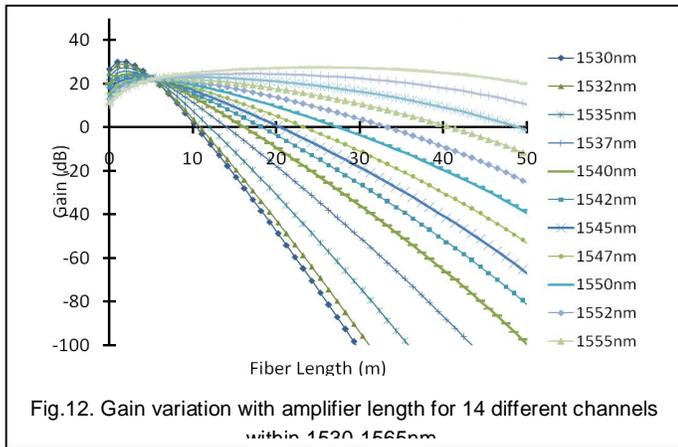


Fig.12. Gain variation with amplifier length for 14 different channels within 1530-1565nm

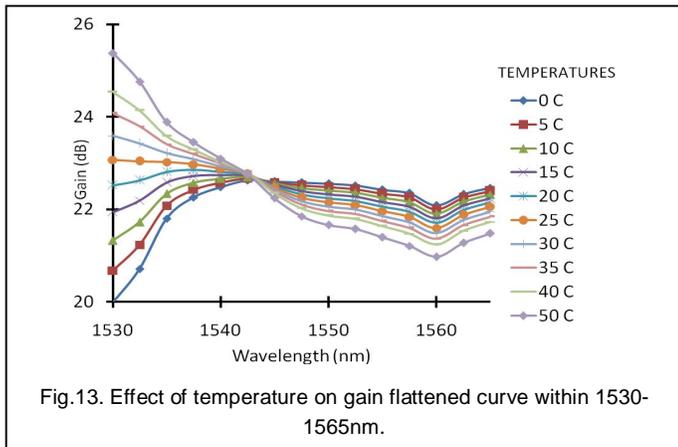


Fig.13. Effect of temperature on gain flattened curve within 1530-1565nm.

TABLE 3

GAIN VARIATION WITH TEMPERATURE

Sl. No.	Temperature in °C	Gain Variation (dB)
1.	0	2.65
2.	5	1.95
3.	10	1.37
4.	15	0.78
5.	20	0.71
6.	25	1.48
7.	30	2.10

Sl. No.	Temperature in °C	Gain Variation (dB)
8.	35	2.64
9.	40	3.30
10.	50	4.40

5 CONCLUSION

The Gain flattening and NF for a bowtie dual stage PM SM EDFA based system depends on pump power, EDF length and operating temperature. The paper has successfully presented the feasibility and limitations of using polarization maintaining single mode bowtie EDF in C-band communication. The strategically positioned equalization filter provided a peak to peak flatness of just 1.1dB and 2.3dB of gain and NF over 35nm range, respectively. However the system faces serious challenges with increasing length with restricted pump power. Temperature analysis of the system, the gain fluctuation has shown its dependence on temperature coefficient of the fiber. The minimum gain variation was obtained at 20°C operative temperature. It puts another limitation to the system, especially with its deployment in extreme geographical areas.

ACKNOWLEDGMENT

The authors would like to thank Prof. Siladitya Sen, Head of the Department, Electronics and Communication Engineering, Heritage Institute of Technology, for his suggestions and support during the work.

REFERENCES

- [1] J. B. Rosolem, A. A. Juriollo and M. A. Romero, "S Band EDFA Using Standard Erbium Doped Fiber," Electronics Letters Oct. Vol. 43 No. 22, (2007).
- [2] S. A. Daud et. Al, "An Experimental Study on S-band Depressed Cladding EDFA", proc. IEEE National Conference on Telecommunication Technologies (2008).
- [3] A.Sano et.al "Ultra-High Capacity WDM Transmission Using Spectrally-Efficient PDM 16-QAM Modulation and C- and Extended L-Band Wideband Optical Amplification," J. Lightwave Technol. 29, p.p.578-586 (2011).
- [4] M. Tiisch, C.A. Hulse, K.D. Hendrix, R.B. Sargent, "Design and demonstration of a thin-film based gain equalization filter for C-band EDFAs", proc. NFOEC conference.
- [5] B. Nagaraju, Pratima Jiaswal, R.K.Varshney, B.P.Pak, "An inherently gain flattened EDFA design based on asymmetric twin core fiber," ICOP-International Conference on Optics and Photonics, Chandigarh, India, 2009.
- [6] R.D Birch et.al., "Fabrication of polarization maintain fibers using gas-phase etching", Electronics Letters, 18(24), p.p.1036-1038 (1982).
- [7] C.R.Giles and E.Desurvire," Modeling of erbium doped fiber amplifiers", J.Lightwave Tech. 9 (2), 271 (1991).
- [8] Y.Tteyam et al, "Correlation between the basicity and optical property of Er³⁺ ion in oxide glasses", Materials Science and Technology, Vol.1 (2006).
- [9] Alan E. Willner, S.M.Hwang," Transmission of many WDM channels through a cascade of EDFA's in Long-Distance links and ring networks", Journal of Lightwave Techn., Vol.13, No.5 (1995).

- [10] V.Sinivasagam, K.Dimyati, R.D.Singh and A.Mustafa, "C-Plus L-band EDFA with over 70nm operating range", IEEE TENCON (2004).
- [11] S.K.liaw et.al., "Investigate C+L band EDFA/Raman Amplifiers by using the same pump lasers", Proc. Of IEEE JCIS (2006).
- [12] D.McCumber," Theory of phonoterminated optical masers," Phys. Rev.,vol.134, p.p. A299-A306 (1964).
- [13] W.J.Lee et.al.," Simulation for the effect of cascaded Gain temperature fluctuations", IEEE CLEOY (1999).

Numerical Simulation of Non-Premixed Turbulent Methane-Air Combustion.

O.A Erinfolami, O.S Ismail, O. Oyewola

Abstract: This paper examined the heat-transfer coefficients and the Nusselt number for turbulent flow for a methane-air flame occurring in cross-flow streams in an open duct. The governing equation of momentum, energy, mass fraction and radical nuclei concentration were solved in a fully coupled fashion at each control volume. The model encompasses conservation equations of gas components participating in the process and energy equation. A FORTRAN code was written with a Tridiagonal matrix solver (TDMA) used to obtain the temperature, velocity and particulates profiles on the flat plate. The Nusselt number for the flow was calculated and plotted as a function of the calculated Rayleigh number. It was observed that a curve obtained with temperature and pressure far from the critical region approaches the line obtained with a classic correlation.

Index Terms— Computation Fluid Dynamics, Eddy dissipation model, Large eddy simulations, Non-Premixed combustion, Probability Density Functions, Reynolds stress model, Reynolds Averaged Navier-Stokes, Turbulence,.

1 INTRODUCTION

Nearly all methods of engineering power production involve fluid flow and heat transfer as essential processes.

The same processes governed the heating and air-conditioning of buildings. Major segments of the chemical and metallurgical industries use components such as furnaces, heat exchangers, condensers, and reactors (nuclear), where thermo-fluid processes are at work. Aircrafts and rockets movements owe much too to fluid flow, heat transfer, and chemical reactions. In the design of electrical machinery and electric circuits, heat transfer is also often the limiting factor. The pollution of the natural environment is largely as a result of heat and mass transfer activities. Similarly are storms, floods, fires and some other natural events. Due to change in weather conditions, the human body resorts to heat and mass transfer for its temperature control.

Combustion is one of the oldest technical processes, and is an everyday phenomenon. It is the dominant mode of energy conversion. Today, combustion related researches are driven mainly by three objectives: Obtaining maximum efficiency, with an eye on the limited supply of fossil fuels; minimizing pollution to preserve the environment and achieving safe and stable operation of machinery and equipments.

The main challenge for combustion researchers, however, is to accomplish all these goals simultaneously. Turbulent non-premixed flames exhibit phenomenon found in other turbulent fluid flows. In some circumstances a thin flame sheet forms a connected but highly wrinkled surface that separates the reactants from the products. This flame surface is convected, bent, and strained by the turbulence and propagates (relative to the fluid) at a speed that can depend on the local conditions (surface curvature, strain rate, etc.). Typically, the specific volume of the products of combustion is seven times that of the reactants, the flame surface being a volume source. Because of this large volume source there is a pressure field associated with the flame surface that affects the velocity field and hence indirectly affects the evolution of the surface itself.

Looking at the detailed structure of a turbulent non-premixed flame, we can examine mean quantities. Within the

flame there is a mean flux of reactants due to the fluctuating components of the velocity field. Contrary to normal expectations and observations in other flows, it is found that this flux transports reactants up the mean-reactants gradient, away from the products (hence counter gradient diffusion).

A second notable phenomenon is the large production of turbulent energy within the flame: Behind the flame the velocity variance can be 20 times its upstream value. Both of these phenomena result from the large density difference between reactants and products and from the pressure field due to volume expansion. There is a wide variety of theories and models for premixed and non-premixed turbulent flames. More ambitious are the probabilistic field theories that attempt to calculate statistical properties of a flame as functions of position and time. In non-premixed flames, also called diffusion flames, the fuel and the oxidizer are mixed during the combustion process itself. Both premixed and non premixed flames can further be classified as either laminar or turbulent depending on the regime of gas flow.

There is an extensive literature describing the development of modeling techniques suitable for large-scale engineering simulations. Early studies of this type using single step chemistry were performed by [1], [2]. More recently, [3] have performed direct numerical simulations of turbulent, premixed hydrogen flames in three dimensions with detailed hydrogen chemistry. Bell et al. [4] have performed similar Numerical Simulation of Premixed Turbulent Methane Combustion. Three different statistical techniques use in turbulent modeling are the Large Eddy Simulations (LES), Reynold Average Navier-Stokes Simulations (RANS) and Probability Density Function (PDF) transport equations model [5]. In LES, the volume-averaging of the governing equations is considered, but will require large computational resources for routine simulations. It is however considered to be a promising tool. RANS technique solves the governing equations by modeling both the large and the small turbulent eddies, taking a time – averaged of the variables being considered. The PDF transport equation model consist of a procedure considering the proba-

bility distribution of the relevant stochastic quantities directly by means of probability density functions.

2.0 MODEL.

2.1 Field Model.

The schematic model of turbulent diffusion flame is as presented in Fig 1. The overall effects of flame in the flow field has to be reduced to generation of heat and the dispersal of heat and burnt particules by either natural air or draughted air towards an object for heating. To carry out detail analysis of actions reflecting occurrences in the flow field a model of the flow field employing time averaged Navier–Stokes equations.

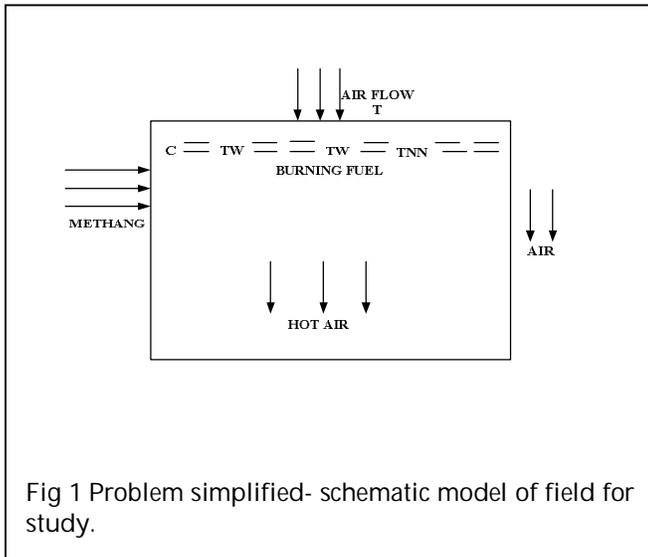


Fig 1 Problem simplified- schematic model of field for study.

2.2 Flame configuration and Field Definition.

For a two-dimensional parabolic flow, when a steady two-dimensional flow has one-way space coordinate, it is called a two dimensional flow. Most flame flow has a predominant velocity in one way coordinate and hence the convection always dominates the diffusion in that coordinate. It is this feature that imparts the one way character to the stream wise direction. No reverse flow in this direction is acceptable. Examples of two-dimensional parabolic flow are plane or axisymmetric cases of boundary layers on walls, duct flows, jets, wakes, and mixing layers. The solution for such situation is obtained by starting with known distribution of the variable, Φ , at upstream station and marching in the streamwise direction. For every forward step, the distribution of Φ in the cross stream coordinate is calculated at on streamwise station. Thus computationally only one-dimensional problem needs to be handled, for which TDMA can be employed as solvers to the discretized equations of these variables Φ .

2.3 Governing Equations.

Employing a conservative finite – volume methodology requires all the governing equations to be expressed in a conservative form in which tensor notation is generally employed. The basic governing equations are for the conservation of mass, momentum and energy.

2.3.1 Continuity equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_j} (\rho u_j) = 0 \quad (1)$$

Where u_j is the j^{th} Cartesian component of velocity and ρ is the fluid density. ($j = 1,2,3$)

2.3.2 Momentum equation

$$\frac{\partial}{\partial t} (\rho u) + \frac{\partial}{\partial x_i} (\rho u_i u_i) = \frac{\partial p}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_i} + \rho f_j \quad (2)$$

Where p is the static pressure, τ_{ij} is the viscous stress tensor and f_j is the body force. For Newtonian fluids τ_{ij} can be expressed as:

$$\tau_{ij} = \mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) - \frac{2\mu}{3} \left[\frac{\partial u_k}{\partial x_k} \right] \delta_{ij} \quad (3)$$

Where μ is the fluid dynamic viscosity and δ_{ij} is the Kronecker delta

2.3.3 Energy Equation

The equation for the conservation of energy can take several forms. The static enthalpy form of the energy equation can be expressed as:

$$\frac{\partial}{\partial t} (\rho h) + \frac{\partial}{\partial x_j} (\rho u_j h) = -\frac{\partial q_j}{\partial x_j} + \frac{\partial p}{\partial t} + u_j \frac{\partial p}{\partial x_j} + \tau_{ij} \frac{\partial u_i}{\partial x_j} - \frac{\partial}{\partial x_j} (J_{mj} h_m) \quad (4)$$

Where J_{mj} is the total (concentration – driven + temperature – driven) diffusive mass flux for species m , h_m represents the enthalpy for species m , and q_j is the j – component of the heat flux. J_{mj} , h_m and h are given as:

$$J_{ij} = -\rho D \frac{\partial Y_m}{\partial X_j} \quad (5)$$

$$h_m = \int_{T_o}^T C_p m (T) dT + h^o f_m \quad (6)$$

$$h = \sum_{m=1}^n Y_m h_m \quad (7)$$

D is the diffusion coefficient, C_p is the constant – pressure specific heat, and h_f^o is the enthalpy of formation at standard conditions ($P_o = 1.atm., T_o = 298K$).

The Fourier’s law is employed for the heat flux:

$$q_j^n = -K \frac{\partial T}{\partial x_j} \quad (8)$$

2.3.4 Mixture Fraction

$$\frac{\partial}{\partial t} (\rho f_k) + \frac{\partial}{\partial x_j} (\rho u_j f_k) = \frac{\partial}{\partial x_j} \left(D \frac{\partial f_k}{\partial x_j} \right) \quad (9)$$

Where D is the diffusion coefficient, f_k is the mixture fraction for the k^{th} mixture.

2.3.5 Standard K - ε model expressed

$$\frac{\partial}{\partial t}(\ell k) + \frac{\partial}{\partial x_j}(\ell^{u_j k}) = \frac{\partial}{\partial x_j} \left(\Gamma_\phi \frac{\partial k}{\partial x_j} \right) + P_k - \ell \varepsilon \quad (10)$$

$$\frac{\partial}{\partial t}(\ell \varepsilon) + \frac{\partial}{\partial x_j}(\ell^{u_j \varepsilon}) = \frac{\partial}{\partial x_j} \left(\Gamma_\varepsilon \frac{\partial \varepsilon}{\partial x_j} \right) + C_1 \frac{\varepsilon}{k} P_k - C_2 \ell \frac{\varepsilon^2}{k} \quad (11)$$

Here $P_k = T_{ij} \left(\frac{\partial u_i}{\partial x_j} \right) = \mu_T S^2$ (12)

$S =$ mean strain rate $= 2S_{ij} S_{ij}$ (13)

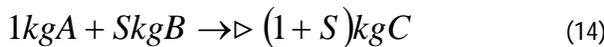
P_k – generation of turbulent due to buoyancy

2.3.6 Chemistry /Reaction Model

The reaction model employed here for CFD development is the instantaneous chemistry model in which the reactants are assumed to react completely upon contact. The reaction rate is infinitely rapid and one reaction step is assumed. Two reactants, which are commonly referred to as “fuel” and “oxidizer”, will be involved. A surface “flame sheet” separates the two reactants (this assumption can be made only for non-premixed flame).

The eddy dissipation model for the chemistry of reactions, Standard EDC model

For combustion reactions, a single step irreversible reaction with finite reaction rates



S is the stoichiometric mass requirement of species B to consume 1kg of A.

Mixture composition is determined by solving for only two variables, mass fraction of specie A, Y_A and mixture fraction ε_j . Using (13), above

$$\frac{\partial}{\partial t}(\ell Y_A) + \frac{\partial}{\partial x_j}(\ell^{u_j Y_A}) = \frac{\partial}{\partial x_j} \left(\frac{\mu_T}{Sc_T} \cdot \frac{\partial Y_A}{\partial x_j} \right) + R_A \quad (15)$$

and

$$\frac{\partial}{\partial t}(\ell Y_A) + \frac{\partial}{\partial x_j}(\ell^{u_j \varepsilon_j}) = \frac{\partial}{\partial x_j} \left(\frac{\mu_T}{Sc_T} \cdot \frac{\partial \varepsilon}{\partial x_j} \right) \quad (16)$$

Where R_A – tone averaged reaction rate. Sc_T - turbulent Schmidt number, validity requires Sc_T to be same for all species.

β - conserved combined variable

$$\varepsilon = \frac{\beta - \beta_\infty}{\beta_0 - \beta_\infty} \quad (17)$$

For components A and B

$$\beta = Y_A - \frac{Y_B}{S} \quad (18)$$

$\beta_\infty \equiv$ value of β at a B rich inlet, $\beta_0 \equiv$ value of β at a A rich inlet

For species A, reaction rate is expressed as

$$R_A = A \cdot \ell \cdot \frac{\varepsilon}{k} \cdot \min \left(Y_A, \frac{Y_B}{S}, B \frac{Y_C}{1+S} \right) \quad (19)$$

$S \equiv$ mixture wt of components B, The constants $A = 4.0$, $B = 0.5$, for micro mixing requirements

$$\Gamma_{m12} = \frac{3}{2} \cdot \frac{k}{\varepsilon} + \frac{1}{2} \ln(Sc) \left(\frac{Y}{\varepsilon} \right)^{\frac{1}{2}} - 5.53 \quad (20)$$

Further work (comprehensive energy calculations) Num flame

Total Energy in a chemical reaction is given by

$$E = h - \frac{P}{\ell} + \frac{V^2}{2} - h \equiv \text{sensible enthalpy} \quad (21)$$

$$h = \sum Y_i h_i + \frac{P}{\ell} \quad \text{Species } i, \quad h_i = \int_T^T c_{p_i} dT$$

one step reaction model



Methane activation energy $E = 48.0$ cal/mole

A = the pre exponential factor = 8.3×10^{16} , h_j – enthalpy of component j in a mixture, C_p of mixture with temp variations, ℓm - density of mixture

2.3.7 Heat and Mass Transfer Model

Transport Equation for energy can be written as

$$\frac{\partial}{\partial t}(\ell E) + \frac{\partial}{\partial x_j}[\mu_i(\ell E + P)] = \frac{\partial}{\partial x_j} \left(k_{eff} \frac{\partial T}{\partial x_j} \right) + S_h \quad (22)$$

Where E = total energy, K_{eff} = effective thermal conductivity

$$K_{eff} = K + \frac{C_p \mu_t}{Pr_t} \quad (23)$$

K – thermal conductivity of component, Turbulent diffusion coefficient, $\Gamma_t = \frac{\mu_t}{Sc_t}$

Effective diffusion coefficient, $\Gamma_{eff} = \Gamma_\phi + \Gamma_t$

3.0 THE SOLUTION PROCEDURE.

3.1 Development Of The Simpler Algorithm.

Once Navier – Stokes partial differential equations governing a flow field are formulated it is incumbent to develop the means of solving the equation. One of the popular methods in use is the method of procuring numerical solution via the SIMPLE algorithm. The procedure is as follows.

If ϕ is a variable and represented by a polynomial of the form, in x-direction

$$\phi = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m \quad (24)$$

The unknowns $a_0, a_1, a_2, \dots, a_m$ are calculated by treating the dependent variable x at a finite number of locations (the boundaries) on a grid of a chosen domain. The methods involves the generation of a set of algebraic equations for these unknowns and the prescription of an algorithm for solving the equation.

3.2 Control Volume Formulation.

First we replace a given continuum with a grid formation of continuous discrete units over the given space, the flow field. The calculation domain is divided into a number of non – overlapping control volumes such that there is one control volume surrounding each grid point. The differential equation is integrated over each control volume. Piecewise profiles expressing the variation of ϕ between the grid points are used to evaluate the required integrals. The result is the discretization equation containing the values of ϕ for a group of grid points. It is not compulsory that distances between the grids are equal. The use of non – uniform grid spacing is often desirable, to deploy computer power effectively, we should note that close to – accurate experimental results are obtained when grid spacing are fine. The number of grid points that should be distributed and numbers depend on the nature of problems to be solved.

3.2.1 Boundary Condition.

The boundary conditions for the two dimensional flame modeling described below:

- (1) Symmetric centerline
- (2) Outlet: The outlet is modeled by means of a pressure outlet; this pressure is set to be atmospheric.
- (3) Outlet or burner wall: Depending on the choice, this boundary is an outlet or a wall. When a free flame is modeled, this is a pressure outlet. When a flame in a box is modeled a burner wall is chosen.
- (4) Burner wall: The burner wall is set to a constant temperature, and a non-slip condition is used.
- (5) Inlet: The inlet will be given by a velocity profile, using the determined profiles from the Fig 1.

3.2.2 Discretization for Two Dimensions.

We shall assume the heat flux q obtained prevails over face area $\Delta y \times 1$. Using similar assumptions for other faces, a differential equation of the form

$$\rho c \frac{\delta T}{\delta t} = \frac{\delta}{\delta x} \left(k \frac{\delta T}{\delta x} \right) + \frac{\delta}{\delta y} \left(k \frac{\delta T}{\delta y} \right) + S \quad (25)$$

can be derived. This can be discretized, as

$$a_p T_p = a_E T_E + a_w T_w + a_N T_N + a_s T_s + b \quad (26)$$

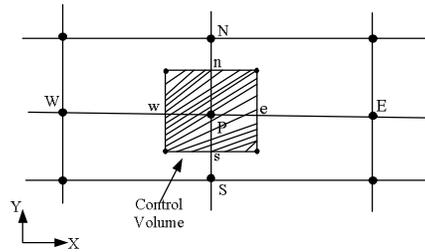


Fig. 2 – Two Dimensional Control Volume

Where

$$\left. \begin{aligned} a_E &= \frac{ke\Delta y}{(\delta x)_e}, a_w = \frac{k_w \Delta y}{(\delta x)_w} \\ a_N &= \frac{k_n \Delta x}{(\delta y)_N}, a_s = \frac{k_s \Delta x}{(\delta y)_s} \\ a_p^o &= \rho \frac{c\Delta x\Delta y}{\Delta t}, b = Sc\Delta x\Delta y + a_p^o T_p^o \\ a_p &= Q_E + a_w + a_N + a_s + a_p^o - S_p \Delta x\Delta y \end{aligned} \right\} \quad (27)$$

3.2.3 Method of Solution of the Linear Algebraic Equations

The solution of the discretization equation can be obtained by the Gaussian – Siedel iteration method. In a system with grids 1,2,3,....., N with 1 and N being the boundary points, the discretization equation can be written as, for a variable T

$$a_i T_i = b_i T_{i+i} + C_i T_{i-i} + d_i \quad (28)$$

Showing Temp T_i is related to temperature T_{i+i} and T_{i-i} . Effort is made to produce a matrix form of the derived equations of any of the given variable and the matrix is subsequently solved using a special form of Gauss-siedel iterative procedures called Tridiagonal algorithm. The method gives better results with one and two dimensional problems. Though convergence of the method is slow for large numbers of grid points, the accuracy could be very high. In order to speed up or slow down the changes from iteration to iteration in order to obtain quick convergence to solution, we employ a overrelaxation as speed up, underrelaxation as slow down. Use of underrelaxation is more common to avoid divergence. Overrelaxation is more employed in Gauss – Siedel procedure.

4.0 NUMERICAL SOLUTION

4.1 Inlet Data.

Methane is delivered through a horizontal orifice that did not protrude into the flame control volume. Composition of gas is as tabulated below.

Table 1 Particulates Data.

No	Property Description	Methane	Ethane	Oxygen	Carbon-dioxide	Water Vapour	Nitrogen
1.	Molecular Wt (Kg)	16.04	30.07	32.00	44.01	36.04	28.02
2.	Mass Fraction (%)	0.86	0.1389	0.2315	0.25	0.45	0.7556

It is assumed that a fuel of 86 % methane content delivered at about 15 m/s at 25° C /1bar via a 24'' pipe (alternatively use mass flow rate for fuel estimates). Air current is in the range 3 – 10m/s at normal 1 bar and 27°C chosen for the study. One step reaction is assumed outward boundary condition applies to flows

4.2. Combustion model

The 2-D flow and mixing fields were resolved by the solution of the 2-D, axisymmetric forms of the density-weighted fluid flow equations, supplemented with the k-ε model. Transport equations of several species are solved with equations for the mixture fraction and its variance. Solution of the transport equations is achieved using backward difference for first derivatives centered differencing for second derivatives with the upwind scheme for flow constants as developed by [6]. The solver employed is the one-way adaptive TDMA procedure as stated above.

The conservation equations of species mass fraction are transformed into the flamelet space. The flamelet equations are solved in pre-processing. The stationary solution is stored in tables containing the profiles of temperature and mass fractions for all chemical species as function of the mixture fraction, its variance and the scalar dissipation rate. The coupling of chemistry and flow field is performed via the mixture fraction, its variance and the scalar dissipation which were provided from the flow field calculations. These values at each computational cell are used to extract mean scalar properties from the chemistry look up tables. The flow field properties are updated and iterations continue until convergence criteria are met. The governing equations are discretized using the finite volume method in an axisymmetric Cartesian coordinates. The Revised Semi-Implicit Method for Pressure Linked Equation, SIMPLER numerical scheme is used to handle the pressure and velocity coupling. The computational domain covers an area from 0 to 1.0 m in the axial direction and 0 to 0.40m in the radial direction. Totally 62(x) × 24(y) uniform grids are used in the simulations. It has been checked that the further increase of grid number does not significantly influence the simulation results. The mass flow rate, total tem-

perature, turbulence intensity and hydraulic diameter are specified for the inlet boundary. At the outlet region, outflow condition is assumed and the symmetry condition on the side boundary. Therefore each unknown will appear in more than one equation in the system, and these equations must be solved simultaneously to give all unknown quantities.

4.3 Grid Size and Mesh generation

A uniform size adaptive grid system is employed. Grid testing was carried out on a limited scale for different plate lengths and a grid size, the results was satisfactory as changes in results conform to expectations. It was observed that as grid size numbers increases the number of iterations required for convergence also increases considerably and beyond certain limit convergence to a solution was not achieved.

4.4 Boundary Conditions

Boundary conditions at the inlet are given separately for the fuel stream at the central jet and the air stream cross-flow. The streams are considered to enter the computational domain as plug flow, with velocities calculated from their respective flow rates. The temperatures of fuel and air are specified. In conformation with the conditions used the fuel flow rate is taken as 0.9517 kg/s and the air flow rate is calculated to be 3.7638 kg/s. The temperatures for both the streams at inlet are 300 K. The temperature for air is taken to be 300 K. The inlet velocities were calculated from the respective mass flow rates and densities of air and methane fuel using mass transport equations. No soot is considered to enter with the flow through the inlet plane. Considering the length of the computational domain to be 1.0m x 0.4 m, the fully developed boundary conditions for the variables are considered at the outlet. In case of reverse flow at the outlet plane, which occurs in the case of buoyant flame, the stream coming in from the outside is considered to be atmospheric air. Axi-symmetric condition was considered at the central axis, while at the wall a no-slip, adiabatic and impermeable boundary condition is adopted.

4.5 Stability Criteria and Convergence Conditions

The choice of the incremental time-step was done very carefully as it is directly related to the stability of the explicit scheme. The first criteria demands that pure advection should not convey a fluid element past a cell in one time increment as the difference equations consider fluxes only between adjacent cells. Hence the time increment should satisfy the following inequality. Solution use an under-relaxation factor of 0.2 - 0.5 to achieve quick convergence as suggested by previous scholars. The corrected velocity is employed to determine convergence of derived solution.

5.0 RESULTS AND CONCLUSIONS.

5.1 Results, Observations and Discussions.

The results presented below revealed the events occurring in the visible forms of flame that we see. Flames are known to exhibit wavy pattern of flow due to recirculation currents found within its flow, this is shown by the negative values of particulates as seen in fig 2a and fig. 2b for the concentrates

under consideration.

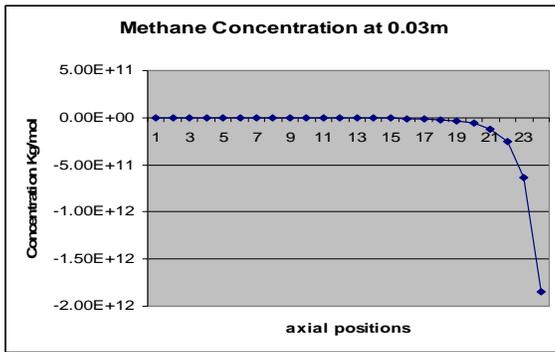


Fig 2a CH₄ particulate concentrations at 0.03m

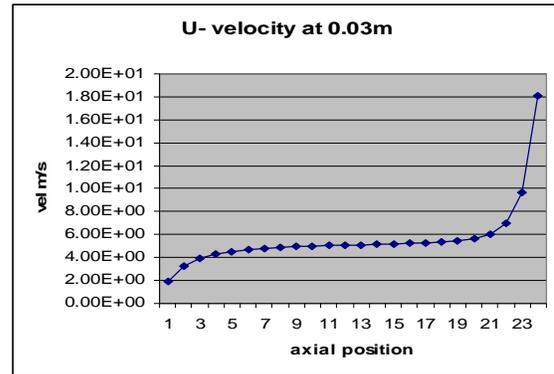


Fig 3b U-velocity distribution at 0.03m

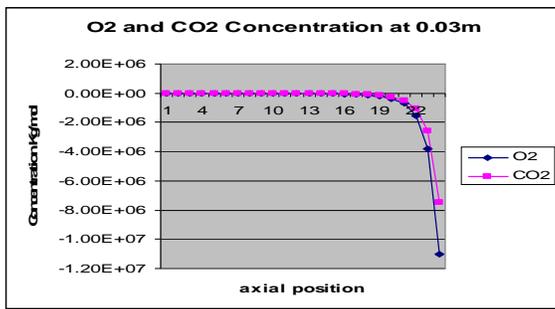


Fig 2b O₂ and CO₂ particulate concentrations at 0.03m

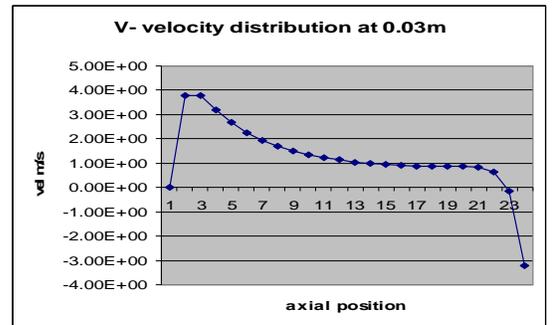


Fig 3c Velocity distribution at 0.03m.

In some circumstances a thin flame sheet forms a connected but highly wrinkled surface that separates the reactants from the products. The flame surface is convected, bent, and strained by the turbulence and propagates (relative to the fluid) at a speed that can depend on the local conditions (surface curvature, strain rate, etc.) Further, the temperature (Fig 3a) shows definite increases as well as the u-velocity (Fig 3b). At this point the v-velocity (Fig 3c) of flow of methane gas/vapour begins to fluctuate rapidly this evidently supportive of buoyancy effects accompanying temperature rise till the methane ignition temperature is reached.

This rapidly changing v-velocity of methane as a result of rapid temperature changes and resultant buoyancy continued beyond 0.1m till about 0.3m where stable positive values were obtained till the end of the control volume (Fig 4a and Fig 4b).

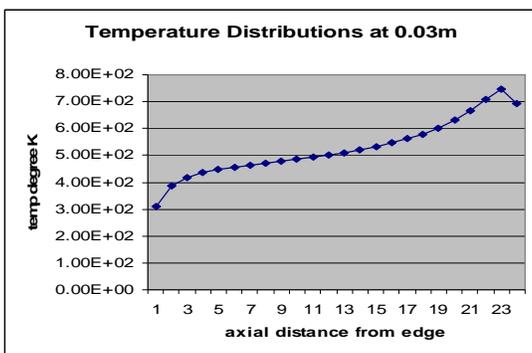


Fig 3c Temperature distribution at 0.03m

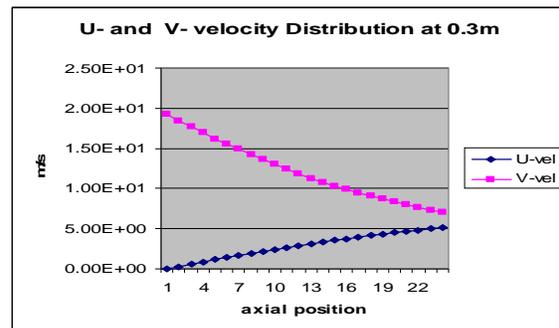


Fig 4a Velocity distribution at 0.3m

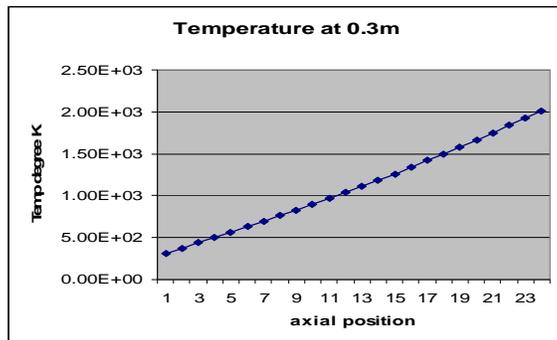


Fig 4b Temperature distribution at 0.3m

One finds that the pollution of the solution with oscillations is dependent on the domination of the convection terms over other terms of the differential equation, in practice most often diffusion terms in the governing transport equations. The higher the Peclet and flow Reynolds numbers are, the more dominant is the convection term and the stronger is the pollution with oscillations.

As the cross wind velocity increases, a transition from buoyancy dominated flow to cross wind dominated flow can be noted along with a reduction in oscillation amplitudes. The changes of stability of flow begins to emerge in the diffusion flame leading to positive values showing that the v-velocity component profile presents in the flow is moving up. This effect is not only attributed to the buoyancy forces, but also due to the presence of the wall normal to the stream-wise direction. It is evident that the buoyancy forces have their impact in this region of the flame.

5.2 Conclusions.

A finite difference scheme for solving the fluid dynamic equations of two dimensional elliptic flows has been used to predict turbulent diffusion jet discharging perpendicularly into an unconfined cross-flow air current. The mathematical model employed a standard k-e model to calculate the distribution of Reynolds stresses. The turbulent gas phase, non-premixed combustion process was model using probability density function approach assuming one step fast chemical reactions. Soot calculation was omitted in this study as its formation was considered negligible. The purpose of the numerical modeling/simulation on the turbulent diffusion flame is to determine the profiles of variables and flow patterns for velocity, species concentration profiles, and temperature in the combustor area.

REFERENCES

- [1] Trouve, A. and Poinso, T., J. Fluid Mech., 278:1-31 (1994).
- [2] Zhang, S. and Rutland, C. J., Combust. Flame, 102:447-461
- [3] Tanahasi, M., Fujimura, M., and Miyauchi, T., Proc. Combust. Inst., 28:529-535 .2000.

[4] Bell, J. B., Day, M. S., and Grcar, J. F., Proc. Combust. Inst., 29 (2002).

[5] Zegers R.P.C., Flame stabilization of oxy-fuel flames, MSc. Thesis, Eindhoven University of Technology. Report number WVT 2007.13

[6] Patankar, S.V., Numerical Heat Transfer and Fluid Flow, Taylor & Francis, Philadelphia / New York.. 1980.

Blind Handicapped Vs. Technology: How do Blind People use Computers?

Prof. Reeta Singh

Abstract— Several companies today market computer programs that allow a blind person to use a standard computer. These computer programs are called "screen readers". A screen reader is itself a standard Windows computer application, but its job is to run alongside the other programs running on a computer and "tell" the blind person what is on the screen. Because a blind person cannot see what is on the screen, a screen reader typically has a built-in speech synthesiser which, although perhaps sounding a bit like a robot, speaks information to the user through the normal sound speakers of the computer itself. People with some limited sight typically use a different kind of screen reader which magnifies and enhances the image on the screen to make it easier to see, and some people use both speech and magnification at the same time. But whether a person uses speech or magnification, typically the screen reader is just a computer application that comes on a CD and easily installs on most computers.

Index Terms— Blind- handicapped, Technology for blind handicapped, Screen Reader, Blind can use computers or not, Disabilities, Internet for Blind, Computer Application for blind.

◆

1 INTRODUCTION

The last decade has seen the triumph of the rich graphical desktop, replete with colourful icons, controls and buttons all around the screen, controlled by the mouse pointer moving about the screen clicking and dragging. This is not, on the face of it, a usable environment for blind people, but use it they must.

The first is to use a **screen reader**. This is an application that attempts to describe to the blind user in speech what the graphical user interface is displaying. It turns the visual output of the standard user interface into a format that is accessible to a blind user. In practice this means driving a Braille output device - a row of Braille cells with mechanical pins that pop up and simulate Braille characters under the user's fingers - or, more commonly, a text-to-speech synthesizer. We will deal exclusively with these text-to-speech users in the rest of this article because they form the great majority of users, actual and potential. The screen reader acts almost as a sighted companion to the blind user, reading out what is happening on the screen - popup boxes, command buttons, menu items, and text. Ultimately screen readers have to access the raw video output from the operating system to the screen and analysing it for information that should be presented to the user. This is a complex process, as you would expect from an application that is attempting to communicate the complicated graphical user interface in a wholly non-visual way. There are many screen readers available, including JAWS from Freedom Scientific, Window Eyes from GW Micro. If you have Windows 2000 or XP, you'll find that Microsoft have included a basic screen reader in the operating system, called Narrator: try activating it, opening Notepad and typing some text or checking your email without looking at your screen.

The goal of a screen reader is to make it appear to the user as if the current application was itself a talking application designed specifically for blind users. This is difficult to accomplish. Applications often have particular user controls or methods of operation that must be supported by the screen

reader. For example, a spreadsheet program operates very differently from an email client. This forces screen reader developers to adapt their programs to support specific applications, typically the market leaders like Microsoft Word. It also means that applications that utilise simple interface components like menus and text boxes will work best with screen readers. Those with non-standard interface components like 3D animations may be difficult for a screen reader to access.

2 SO HOW DOES A BLIND PERSON ACTUALLY KNOW WHAT IS ON THE SCREEN?

People who are totally blind are absolutely not able to interact with the computer without assistive technologies. In order to overcome this barrier, they mostly use **screen reader** software and Braille displays. In simple terms, a screen reader system speaks all the information in a human voice which comes on the screen as well as the text which is typed on the keyboard. A Braille display makes the same information appear on a Braille line which blind people can read with their fingers.

However, a screen reader is much more complicated in practice. It is also important that blind people are able to navigate quickly on the screen and find information as they need it. Therefore, screen reader systems are loaded with functionality which read a portion of the screen according to certain different criteria. The more simple ones would read the current character, the current word or the current line. More complex ones would read the status line of an application, the title bar, a certain window, or the current item on the menu as the user navigates. A Braille display is usually an addition to a screen reader. It is a small unit which lays by the keyboard and displays one line of information in Braille, mostly the same which the screen read announces with speech. This helps blind people understand the layout of the screen better, and read

texts which is more difficult to understand with speech, for example more complex tables, or texts which contain words in more than one language, such as dictionaries.

The effectiveness of a screen reader greatly determines the effectiveness of blind people on the computer. Long ago, screen readers only allowed to read the screen line by line, so people had to hunt for information they needed. Today, practically any piece of information can be assigned with a hotkey. Different hotkeys would announce different information in different applications. For example, one hotkey would announce the misspelled word in a Microsoft Application, another would read the current table cell in Internet Explorer, etc.

Nowadays, there is a larger variety of screen readers available. Some of the most popular ones, which have been around can be very costly, more than \$1000. There are also lower cost screen readers available, and currently free, open source ones are being developed.

Operating systems also include some kind of a screen reader, Windows for example uses Narrator, which is a very simple system, not necessarily sufficient for the complex use of a computer for blind people, but definitely very helpful for smaller tasks. Apple built a very sophisticated screen reader into its operating system. When using Linux, the Gnome Desktop by default also contains a screen reader.

When we look at different types of disabilities, people who are blind probably need the most accommodations in technology in order to use the computer. For certain disabilities, a small software or hardware might do the job, but aside from using a screen reader, blind people also need to rely on developers to have their applications coded in an accessible way.

Certain accessibility issues can be corrected by customizing screen readers, but there are some issues which are difficult, or impossible to overcome.

One of the biggest accessibility issue blind people face today is that images are not described with regular text. Web sites, for example are very difficult to use when image links are not labeled, diagrams are not explained with text, or videos do not provide alternative information.



Probably the second largest challenge is when an application is not usable with the keyboard. Very often, navigation and accessing functionality is directly tied to the mouse and a keyboard equivalent to achieve the same task is not available. The use of the mouse, however, requires site, thus blind people are not able to interact with these applications.

Blind people use the computer keyboard just like anybody else. As a matter of fact, it is not necessary to see the keyboard when typing. The best typists do not look at the keys or their fingers. It definitely takes a learning curve to memorize the keyboard and get up to a certain speed, but it really pays off at the end.



Hear what a screen reader sounds like. Listen to the first two paragraphs of this article in a machine sounding voice, and in a more human sounding voice

The screen reader keeps track of what the computer is doing, and speaks or magnifies the necessary information that a user needs in order to use the computer. When you as a sighted person look at the computer screen, of course you see the whole screen, but instinctively you focus your attention on the bit of the screen that is immediately relevant. A screen reader does the same thing. It does not simply read the whole screen, because that would quickly become tedious. But it monitors the screen and automatically tells the blind user the most important information about what is currently happening.

A totally blind person, someone with no sight at all to read the screen, cannot use a mouse. Most functions on the computer can be operated from the keyboard, though this does vary with different programs, and a totally blind person learns how to access the needed functions by using the keyboard. But in some cases, a good screen reader can even make functions accessible

with special keyboard combinations when typically those functions might only be available with the mouse. A blind person, but they can only achieve this by focusing on some very basic tasks such as limited word processing and note taking.

3 SO DO BLIND PEOPLE NEED SPECIAL COMPUTER APPLICATION?

There are some computer applications specially marketed for use by blind people. These are designed to make using a computer as easy as possible for a blind person, but they can

only achieve this by focusing on some very basic tasks such as limited word processing and note taking.

So most blind people use everyday computer applications such as Word, Internet Explorer and Outlook, and actually can do most of the tasks sighted people can.

4 WHAT CAN BLIND AND VISION IMPAIRED PEOPLE DO USING A COMPUTER?

Blind and vision impaired people do many different things in life, so there is almost no limit to the range of tasks we can do on a computer.

- Reading newspapers on the internet
- Internet banking and shopping
- Chat rooms and voice communication over the internet
- Internet radio stations and entertainment
- Looking up information on the internet
- Email correspondence
- Written correspondence
- Personal notes and keeping track of personal information
- Essays and other work to do with study
- Report writing (for employment, committee work etc.)
- Personal or business accounts (spreadsheet or book keeping applications)



5 WHAT DIFFERENCE DOES THIS MAKE TO A BLIND PERSON?

Nowadays surfing the internet and the use of the internet for such tasks as shopping and banking is still felt by many people to be a bit of a gimmick, the reality for blind people is quite different. For example, internet banking such as via the ASB Bank's FastNet service can be easily used by a blind person with only moderate training. So we can read our own bank statements, pay our own bills, and generally keep tabs on our own personal money without assistance from someone else. This means we can be more independent with our personal finances and avoid the vulnerability that sometimes comes from being completely dependent on someone else.

Blind people are becoming more interested in internet shopping, even for the weekly groceries. Most internet shopping sites can be easily used by blind people even without the pictures. This gives us the opportunity to browse for things in a way that sighted people take for granted.

6. WHAT CAN BLIND PEOPLE NOT DO USING A COMPUTER?

Typically, a totally blind person cannot do anything that is mainly graphical, or which involves interacting with moving images. Screen readers, particularly those that use synthetic speech, cannot deal with graphical applications such as painting and drawing. Nor can they cope with animated applications such as most games.

7 SPECIALIZED TRAINING

So yes there is a lot that we as blind people can do using a computer, but by now you have probably realized that the way we work with a computer is somewhat different from the typical way a sighted person would use one. Because of this, blind people do need, at least initially, specialized training, because in that case the way we interact with the application is quite different than for sighted users, so it is not easy for a sighted person to explain what is happening in a way that will make sense to a blind user.

8 Conclusion

Hence, Technologies has proved that Blind people is not Alone blind people can access computers.

Earlier in the paper, we stated that the potential of the Internet could not be understood without reference to other information sources, disability and the larger frameworks of economy and culture as well as the micro-social context. The importance of understanding the perceptions about issues of potentialities, as well as barriers, is underlined by the fact that, despite the putative benefits of Internet for people with disabilities, their level of access is well below that of people without disabilities

Within the micro-social context, the Internet can mean many things to people with a disability: a luxury, a necessity, a way to participate in the information society, a way to gain access to more information than was previously available, or only one of the many ways of accessing information. It is also seen as a technology which may potentially disadvantage them if they cannot access it. Within this individual context, the lack of fit between the needs of the person with a disability and technology is of less concern than the economic practicalities of affording the equipment in the first place. In this sense the economic framework of our society impinges more than, or perhaps before, the technical-cultural.

In the present consumer society, there appears to be a continuum of participation: those who can afford to participate and do, those who can afford to participate but feel they have no need to, those who cannot afford to participate but would like to, and those who cannot afford to participate and feel they have no need to. The consumer society aims to target those who can afford to participate and this group of people tend to be able-bodied. Technology is aimed at an able-bodied, salaried group of people. People with disabilities tend to fall at the

other end of the continuum: those who cannot afford to participate.

We are also living at a time during which information has become one of the most important commodities. New ways of presenting information which reach wider audiences and are cheaper and more efficient are being encouraged. The Internet is seen as the future of information provision because it has a world-wide audience. However, on a local scale, television, radio, newspapers, and newsletters are still recognized as legitimate forms of information and communication. For people who are blind or visually impaired, as with the rest of society, the Internet is a choice among many other media. Television and radio are very important ways in which they get information. Organizations for the blind are extremely important in disseminating information which is printed and family and friends are, as always, information sources. While a consumer society often stimulates more choice through competition, it also tends to converge on ways of disseminating information which are the cheapest. It is becoming clearer to many (particularly government departments) that it is cheaper and more convenient to provide information to the public via the Internet. If this form of information provision becomes the predominant one, those who have difficulty participating physically, economically, because they have a fear of technology, or lack training opportunities, will be disadvantaged.

In the meantime, there are many options available for getting information and the Internet is not yet a necessity. Whether people are being disadvantaged when they have difficulty in accessing a technology which is not as yet a necessity is a difficult question. It seems pointless to answer this question with a blanket statement of 'yes' or 'no', as some people clearly get all the information they need for their lives without using the Internet, and others indicate that they would clearly benefit from using the Internet. It is best to answer this question from the standpoint of each individual's need. As people's needs change according to their changes in lifestyles and life stages, their need to access the Internet likely to change. What is clear is that the choice for participation should be available to everyone.



**"Always in the dark,
never able to see beyond,
but they have a faint glow of hope that will spark,
when they hear you and respond.**

**You can give them hope,
and you can help them strive,
when they can't get pass the challenges you are there to help
them cope,
you are what makes them come alive."**

REFERENCES

- [1] American Foundation for the Blind. (2000). *Optical character recognition systems*. Retrieved June 22, 2001 from the World Wide Web: http://www.afb.org/info_document_view.asp?documentid=1283
- [2] Penney and Associates, Ltd. (1996). *Internet access for people with disabilities, final report*. Retrieved July 21, 2000 from the World Wide Web: <http://indie.ca/strategy/accessreport.html>
- [3] Adaptive Technology Resource Centre. (2000). *Technical glossary*. Toronto, Canada, University of Toronto. Retrieved from the World Wide Web: <http://www.utoronto.ca/atrc/reference/tech/scread.html>
- [4] <http://jcmc.indiana.edu/vol17/issue1/williamson.html>
- [5] www.evengrounds.com/blog/do-blind-people-use-special-keyboard
- [6] www.abcnz.org.nz/download.php?...blind-people-use-computers...
- [7] <http://www.webbie.org.uk/webbie.htm>

Author Detail:-Prof.Reeta Singh, BE Computer, Asst.Prof. At IMCOST College, Thane. Email-Id:- Reeta_singh2007@yahoo.com

Design and Implementation of Thin-filmed Piezoelectric Pressure Sensor

Amira Mahmoud Olayan, Amal Zaki, Hazem Hassan

Abstract— this paper offers an analytical modeling of thin-filmed, multi-layer piezoelectric sensor. In this work, analysis and simulation of a thin plate pressure sensor is carried on using both mathematical models and finite element analyses. Experimental methods, in the form of full-field-of-view laser microscopy, are applied to show the internal structure of the fabricated pressure sensor and update the computational models. The combined approach of the two methods is used to determine the elastic modulus and the thickness of the diaphragm of the pressure sensor. MEMS pressure sensors are based on a piezoelectric capacitive thin diaphragm that is fabricated using standard Integrated Circuit (IC) procedures including anisotropic etching. During operation, the thin diaphragm of the sensor deflects under pressure loadings, which produces electrical outputs. Analytical simulation using COMSOL software and theoretical computations using equations will be offered in order to determine the parameters for optimal design setting.

Index Terms— COMSOL modeling and simulation, MEMS, Piezoelectric microphone, square diaphragm, ZnO.

1 Introduction

There has been significant advances in the area of thin film, multi-layer piezoelectric with application to MEMS. This paper describes a piezoelectric sensor built on a square silicon diaphragms 25mm thick (SD) Using MEMS

2 Theoretical considerations

In this paper, two methods are used to expose the process that determines electro mechanical parameters of multi-layer sensors. The first method uses mathematical formulation to analyze the relationship between applied surface contact pressure on the surface of the sensor and the resulting bending and displacement. Furthermore, the

Technology. The advantages of the new piezoelectric are high sensitive, wide frequency response range, low electrical power consumption, high precision, and simplified instrumentation.

relationship between displacement and voltage is established.

The second method is on the basis of analytical simulation of micro cantilever using COMSOL software, also by using the MATLAB this relation was plotted This Sensor is used to analyze, measure, and expose the pressure.

2.1 Analytical method

The diaphragm of the sensor is modeled using Kirchhoff's thin plate theory [5]. For a fully constrained plate subjected to a uniform pressure, Timoshenko's solution indicates that the maximum deformation, w_{\max} of the diaphragm is [5],

$$w_{\max} = -\beta \frac{Pa^4}{Eh^3} \quad (1)$$

- Author name is Amira Mahmoud Olayan currently pursuing masters degree program in electric engineering in University of Arab Academy for Science, Technology and Maritime Transport , Egypt, 02-01001088809. E-mail: amira.olayan@gmail.com
- Co-Author name is, Amal Zaki in National Authority for Remote Sensing and Space Sciences, Egypt, E-mail: amalzaki@gmail.com
- Co-Author name is Hazem Hassan in University of Arab Academy for Science, Technology and Maritime Transport ,Egypt, E-mail: hazemhali@gmail.com
-

Where, p is the applied pressure, and a , E , and h are the side length, the elastic modulus, and thickness of the diaphragm, respectively. $\beta = 0.0138$ for a square diaphragm.

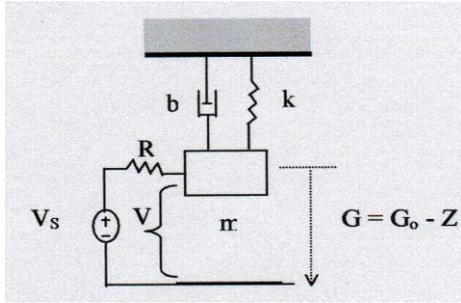


Figure 1: Schematic diagram of the parallel plate model

The voltage equations for the circuit of Figure 1:

$$V_s = iR + V, \quad i = \frac{dQ}{dT} \therefore \frac{V_s}{R} = \frac{V}{R} + i$$

$$\frac{V_s}{R} = \frac{V}{R} + \frac{dQ}{dT} \quad (2)$$

The system mechanical equations can be expressed as following:

$$m \frac{d^2Z}{dT^2} + b \frac{dZ}{dT} + KZ = F_{\text{elect}} \quad (3)$$

And

$$F_{\text{elect}} = \frac{Q^2}{2C_0G_0} = \frac{\epsilon_0AV^2}{2Z^2} \quad (4)$$

In order to examine the energy storage and the power dissipation in this system, it is useful to normalize these equations. Distance is normalized to the initial gap G_0 , time is normalized to the inverse of the un-damped mechanical resonance frequency (ω_0 , voltage is normalized to the pull-in voltage V_{PI} , and charge is normalized to the "pull-in charge" Q_P which is the charge stored when the capacitor voltage is equal to V_{PI} and gap is reduced to 2/3 of the initial gap; The new normalized variables are:

$$z = \frac{Z}{G_0}, \quad t = \omega_0 T, \quad q = \frac{Q}{Q_P}$$

$$v = \frac{V}{V_{PI}}, \quad v_s = \frac{V_s}{V_{PI}}$$

3 Simulation and design

I Computational analyses

The MATLAB code was used as an analyses method, figure 2. Shows the code used and the parameters, where the result was used for the analyzing the relation between the different applied pressure and the deformation (plate displacement) accurse as shown in figure 3:

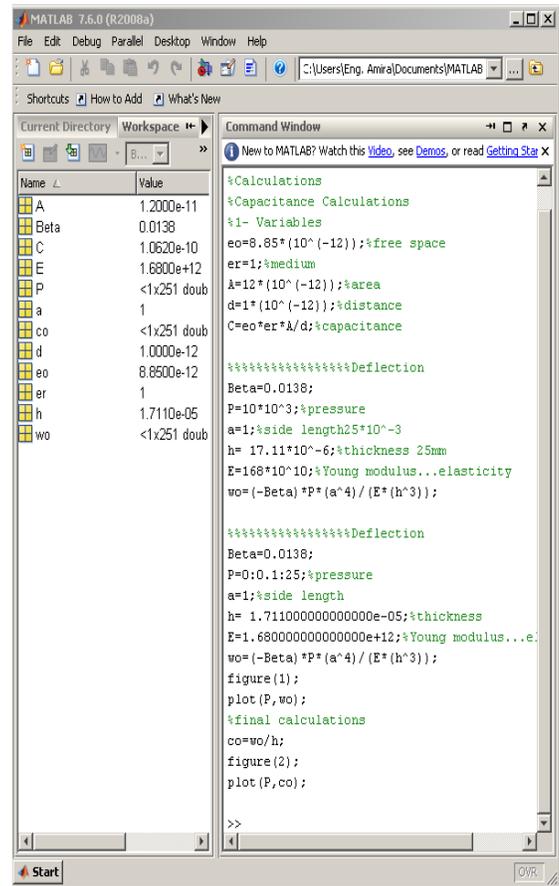


Figure 2: MATLAB code

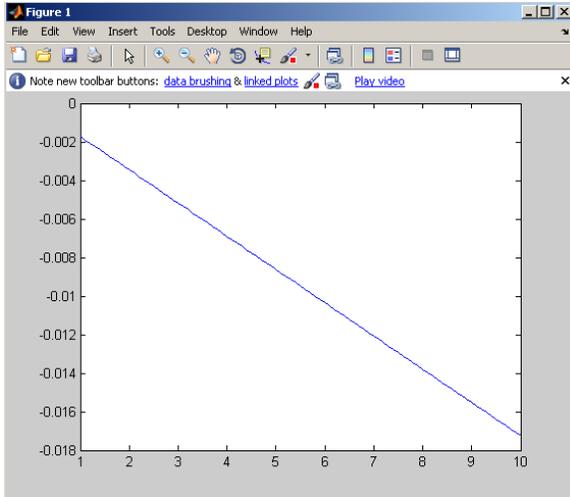


Figure 3: Displacement versus stress applied

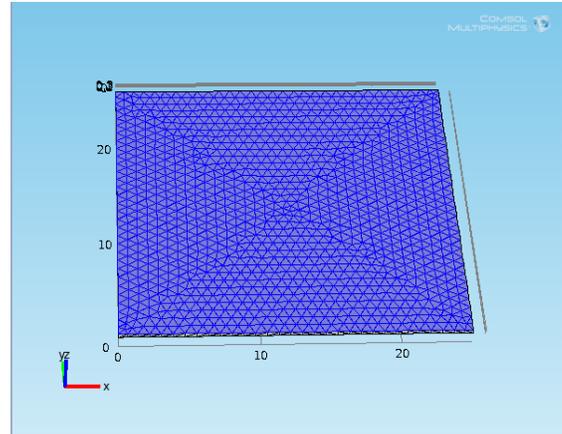


Figure5: the Meshing of the diaphragm layer

II- FEM analyses (COMSOL)

In this work the MEMS pressure sensor is modeled using FEM technique by COMSOL in order to obtain the maximum deflection occurred at the center of the thin plate structure layers, and the distributed voltage through the sensor due to the distributed pressure in the surface area of the thin plate layers, Figure 4 shows representative calculated deformation for a diaphragm subjected to different loading conditions.

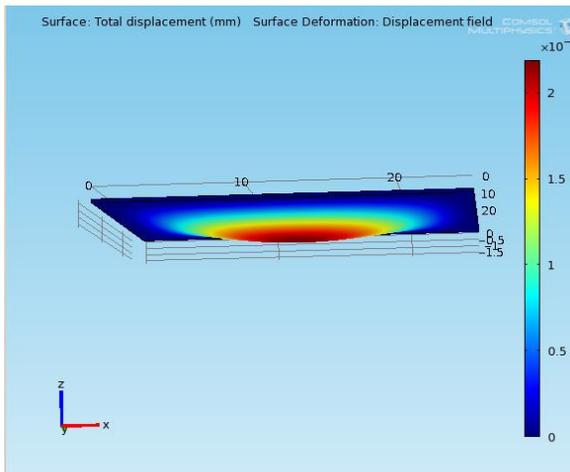
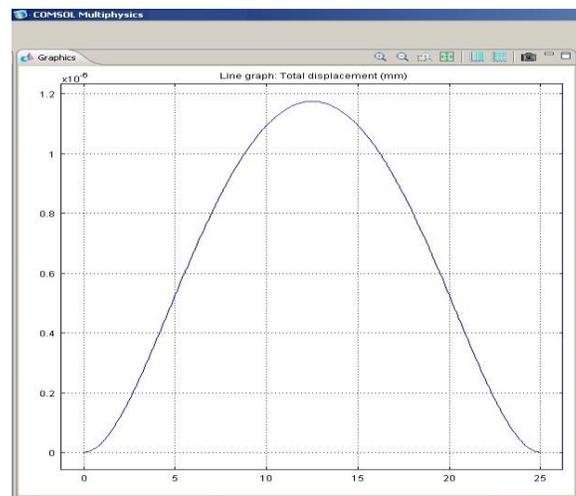


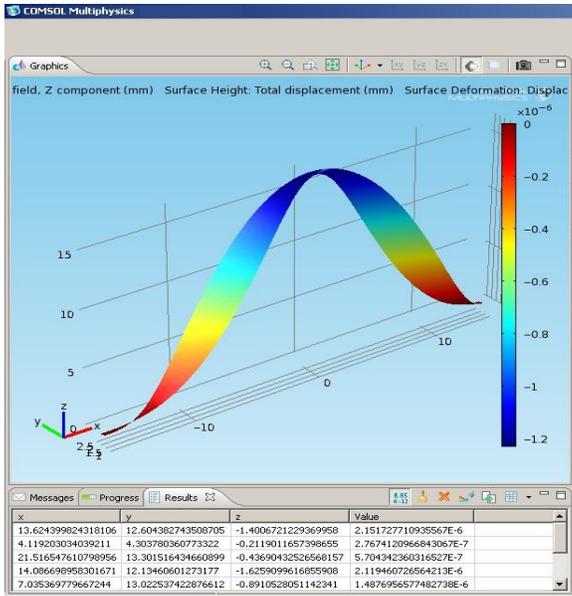
Figure 4: Representative Diaphragm deformation

The meshing procedure of the multilayer diaphragm is shown in Figure 5.

The maximum deflection of the clamped-clamped diaphragm was obtained by applying pressure 0.2bar which gives 1.55um as shown in figure 6 (a), and the diaphragm voltage distribution is shown in figure 6 (b) which show that the maximum surface displacement value is at the center of the diaphragm and reduced at the tip of the diaphragm, and this displacement increased by increasing the applied stress which reach the maximum value at 2 bar then it is saturated because of the ZnO material.



(a) Deflection of the diaphragm



(b) Diaphragm surface height total displacement

Figure 6: Diaphragm simulation output RESULTS

The maximum deformations as a function of applied pressure data obtained with analytical, computational, and experimental results. From these data, the elastic modulus and the thickness of the diaphragm of the MEMS pressure sensor of interest are determined.

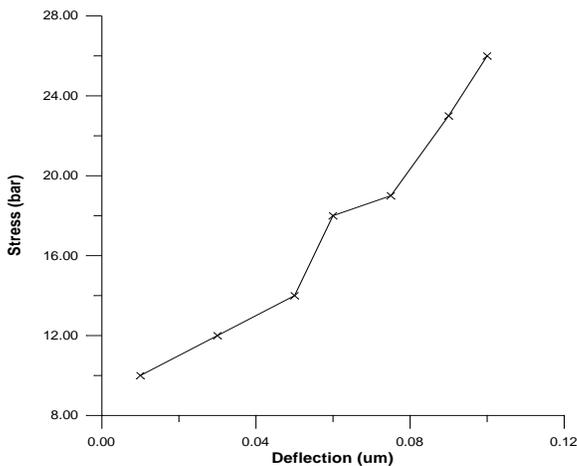


Figure 8: The relation between stress and deflection

A model describing the deflection of a piezoelectric thin plate structure can be derived by appealing to the basic mechanics principles of the static equilibrium and strain compatibility between successive layers in the device. An electromechanical simulator based on FEM technique is used to obtain the relation between the stress and deflection at different operating force. Fig.8 represents this relationship.

4 CONCLUSIONS

The measured capacitance value of 115-120pF on the central region is carried out using the C-V Plotter Instrument. The acoustic sensor showed a flat frequency response from 31.5Hz to 8 KHz. The measured average sensitivity is 50mVrms /Pa and a reasonably linear output over 110-160dB of SPL.

In this work the MEMS acoustic sensor is modeled using FEM technique by COMSOL in order to obtain the maximum deflection occurred at the tip of the thin plate structure layers, and the distributed voltage through the sensor due to the distributed pressure in the surface area of the thin plate layers.

The maximum deflection of the clamped-clamped diaphragm was obtained by applying pressure 0.2bar which gives 30um, and the diaphragm voltage distribution which show that the maximum voltage value is at the centre of the diaphragm and reduced at the tip of the diaphragm, and this voltage increased by increasing the applied stress which reach the maximum value at 2 bar then it is saturated because of the ZnO material.

ACKNOWLEDGMENT

No words would convey the sincerity of my gratitude to GOD followed by the people who have supported me from the start till the end and continue to support and encourage me unconditionally while expecting nothing in return.

A special word of gratitude should be dedicated to my parents and sisters & husband for their unfailing encouragement and moral support. Without them, the task of completing this project would be impossible.

First and foremost I would like to thank Prof. Dr. Hazem Hassan for his invaluable guidance, support and motivation throughout the duration of this research.

I am very thankful to Prof. Dr. Amal Zaki who guided me and has supervised my thesis and was a source of great knowledge that aided in understanding and completing this thesis.

I strongly owe complete appreciation to Eng. Hesham H. Gaber, he was always there for me, supporting, guiding and advising and has he been a mentor to me.

Not to mention everyone in my life, who directly or indirectly had an impact on me and helped to become who I am

REFERENCES

- [1] Antonio, J., Analytical and experimental investigations of high-resolution MEMS pressure sensors, Major\ qualifying project, Mechanical Engineering Department, Worcester Polytechnic Institute, Worcester MA, 2006
- [2] Pryputniewicz, D. R., ACES approach to the development of micro-components, M.S. Thesis, Worcester Polytechnic Institute, Worcester MA, 1997
- [3] Hsu T.R., MEMS & Microsystems: design and manufacture, McGraw-Hill, New York, 2002
- [4] Pelesko, J. A. and Bernstein, D. A., Modeling MEMS and NEMS, CRC Press, Boca Raton FL, 2003
- [5] Timoshenko, S. and Woinowsky-Krieger, S., Theory of plates and shells, McGraw-Hill, New York, 1959
- [6] Lawrence K. L., ANSYS Workbench Tutorial, Schroff Development Corp, 2005

[7] Koppers, J. D., Gouverneur, I. M., Rodgers, M. T., Wenger, J. and Furlong, C., "Dynamic characterization of AFM probes by laser Doppler vibrometry and stroboscopic holographic methodologies," Proc. SPIE, San Diego CA, 2006

Enabling Poor People to Overcome Poverty and Innovative Access to Finance for Small and Medium Enterprises in the United Republic of Tanzania by Rotating Savings and Credit Association Method.

A. Ranjith, R.Thandaiah Prabu, P. Albert Shelton

Abstract— The United Republic of Tanzania is an emerging economy with high growth potential. With per capita gross domestic product (GDP) of US\$500, the economy has shown strong and consistent growth in the last two decades the Tanzanian economy went through a period of successful transition in which economic liberalization and institutional reform led to a recovery of GDP growth to more than 7% per year since 2000. However, during the same period, the country has been unable to achieve significant reductions in poverty. Approximately 90 per cent of the United Republic of Tanzania's poor people lives in rural areas. The incidence of poverty varies greatly across the country but is highest among rural families living in arid and semi-arid regions. The main aim of this paper is to enabling poor rural people to overcome poverty and increase their savings by introducing The Rotating Savings and Credit Association (Rosca) method. It plays an important role as a financial intermediary in many parts of developing countries. It flourishes in both urban and rural settings, especially where formal financial institutions seem to fail to meet the needs of a large fraction of the population.

Keywords— Roscas, upatu, bid, poverty, loan, MDG and household

1 INTRODUCTION

In Tanzania, this in 1999 ranked 156 out of 174 countries in the Human Development Index (HDI), 53% of the population is below the age of 18. It is estimated that Tanzania's economy must grow by 7% per annum. 51% of the population lives on less than \$1 a day. About half or 42% of these live in absolute poverty on less than \$0.75 cents a day [12]. A popular and useful definition of a poor person is someone who doesn't have much money. Among academics, and in the aid industry, this definition has gone out of fashion. Poor people can save and want to save, and when they do not save it is because of lack of opportunity rather than lack of capacity. During their lives there are many occasions when they need sums of cash greater than they have to hand, and the only reliable way of getting hold of such sums is by finding some way to build them from their savings. They need these lump sums to meet lifecycle needs, to cope with emergencies, and to grasp opportunities to acquire assets or develop businesses.

The job of financial services for the poor, then, is to provide them with mechanisms to turn savings into lump sums for a wide variety of uses (and not just to run microenterprises). Good financial services for the poor are those that do this job in the safest, most convenient, most flexible and most affordable way. The poor seek to turn their savings into lump sums by finding reliable deposit takers, by seeking advances against future savings (loans), or by setting up devices like savings clubs and ROSCAs [4].

ROSCAs are a means to 'save and borrow' simultaneously. Rotating Savings and Credit Associations can be found all over the world and go by different names in different regions and countries. In broad terms, a ROSCA can be defined as 'a voluntary grouping of individuals who agree to contribute financially at each of a set of uniformly-spaced dates towards the creation of a fund, which will then be allotted in accordance with some prearranged principle to each member of the group in turn' [2]. It is considered one of the best instruments to cater to the needs of the poor.

Upatu are the Tanzania equivalent of the Rotating Savings and Credit Associations (ROSCA). It enables poor people to convert their small savings into lump sums. The concept of upatu originated more than 1000 years ago. Initially it was in the form of an informal association of traders and households within communities, wherein the member contributed some money in return for an accumulated sum at the end of the tenure.

- A. Ranjith & R. Thandaiah Prabu are currently working as a lecturer in school of electronics and telecommunication engineering in St. Joseph College of Engineering and Technology, Tanzania. Mobile: +255 714949729, +255 656023085. E-mail: aranjithece@gmail.com, thandaiah@gmail.com.
- P. Albert Shelton is currently working as a lecturer in Mechanical engineering in St. Joseph College of Engineering and Technology, Tanzania. Mobile +255 652325117, E-mail: albertshelton85@gmail.com

Participation in Upatu was mainly for the purpose of purchasing some property or, in other words, for consumption purposes [3].

ROSCAs are found worldwide (like Asia, Latin America, and the Caribbean Africa) in countries with vastly different levels of economic development [6]. ROSCA participation is high in Africa [5]. These estimated that, in Central African countries, about 20% of household savings are accumulated in informal ROSCAs [15]. A sample of 115 households in central Kenya showed that 45% were participating in a ROSCA [7]. In a sample in urban Zimbabwe 76% of urban market traders participate in a ROSCA; even though 77% of these traders have a banking account [8]. Taiwan with relatively well functioning credit markets as many as 80% of adults is estimated to belong to ROSCA [9]. ROSCA can provide a commitment mechanism that ties participants hands and commits them to saving patterns and sometimes to spending patterns as well [5]. Several commitment devices that villagers in East Africa use to stick to saving plans, including buying a lock box and throwing away the key [1]. ROSCA is one of the vital methods in various countries to reduce poverty and increase saving potential. The main objective of this paper is to reduce the poverty percentage in Tanzania and increase the participant's percentage in ROSCA like the above mentioned countries.

2 PRELIMINARIES

In this section, we first state the number of people in poverty, population and housing census of Tanzania, changes in the poverty head count in selected countries.

2.1 POVERTY STATISTICS 2001 AND 2007

Overall, in the 16 year period between 1991 and 2007, poverty fell by about 5%. From the table1 There are more poor people today than in 2001 While the percent of people living in poverty (i.e. on less than Tsh.500/- per person per day) went down slightly since 2000/1, because the population has increased, the total number of people living below the poverty line increased by 1.3 million in the same period. Tanzania has signed up to the Millennium Development Goals (MDGs). The First MDG commits Tanzania to reduce poverty between 1990 and 2015 by 50%. In 1991/92 poverty was 38% in Tanzania, so the objective is to reduce poverty to 19% by 2015 [10].

TABLE I
NATIONAL BUREAU OF STATISTICS 2001 & 2007

Year	Population (Tanzania main land)	Poverty rate (%)	Number of people in poverty
2001	32.4	35.6	11.5
2007	38.3	33.4	12.8

2.2 POPULATION AND HOUSING CENSUS

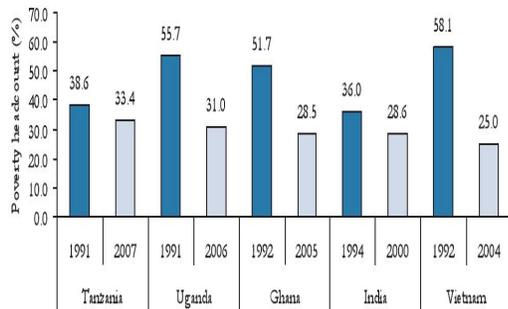
In Tanzania, the latest population and housing census that covered all regions was carried out in 2002. The 2002 census was preceded by three other post-independence population censuses which were conducted in 1967, 1978 and 1988[11]. The above table shows that United Republic of Tanzania, 2002 Population and Housing Census. From Table II Mwanza, Dar es Salaam and Mbeya region has the highest population when compared with other region [11].

TABLE II
UNITED REPUBLIC OF TANZANIA,
2002 POPULATION AND HOUSING CENSUS

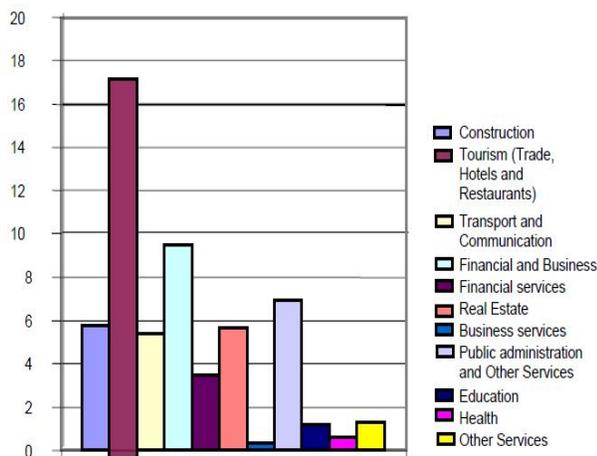
Region	Population(Number)			Households	
	Males	Females	Total	Number	Avg. Size
Country Total	16,910,321	17,658,911	34,569,232	6,996,036	4.9
Dodoma	823,504	875,492	1,698,996	376,530	4.5
Arusha	638,261	654,712	1,292,973	286,579	4.5
Kilimanjaro	667,865	713,284	1,381,149	297,439	4.6
Tanga	797,240	844,775	1,642,015	356,993	4.6
Morogoro	877,238	882,571	1,759,809	385,260	4.6
Coast	440,161	448,993	889,154	200,919	4.4
Dar Es Salaam	1,261,077	1,236,83	2,497,940	596,264	4.2
Lindi	381,359	409,947	791,306	190,761	4.1
Mtwara	534,402	594,121	1,128,523	293,908	3.8
Ruvuma	545,547	571,619	1,117,166	232,340	4.8
Iringa	708,927	786,406	1,495,333	346,815	4.3
Mbeya	990,825	1,079,221	2,070,046	491,929	4.2
Singida	531,015	559,743	1,090,758	217,572	5.0
Tabora	846,196	871,712	1,717,908	291,369	5.9
Rukwa	559,120	582,623	1,141,743	222,868	5.1
Kigoma	807,859	871,250	1,679,109	242,533	6.9
Shinyanga	1,369,581	1,435,999	2,805,580	445,020	6.3
Kagera	999,941	1,033,947	2,033,888	394,128	5.2
Mwanza	1,459,570	1,482,578	2,942,148	495,400	5.9
Mara	653,449	715,153	1,368,602	246,600	5.5
Manyara	534,565	505,896	1,040,461	199,860	5.2
Zanzibar North	67,093	69,860	136,953	27,854	4.9
Zanzibar South	47,830	46,674	94,504	19,937	4.7
Zanzibar Town/West	190,937	200,065	391,002	74,363	5.3
Pemba North	90,943	95,070	186,013	33,019	5.6
Pemba South	85,816	90,337	176,153	29,776	5.9

2.3 POVERTY IN SELECTED COUNTRIES

From the figure, Tanzania's performance in poverty reduction compares poorly relative to comparator countries which like Tanzania were relatively stable and who underwent macro-economic reform in the region (Ghana, Uganda) and in Asia (Vietnam and India). Whereas in Tanzania headcount poverty declined by 2.4% between 1991 and 2007, it dropped in Uganda, Ghana and Vietnam by 10 times as much: approximately 23 to 24%. India too achieved a much higher reduction in poverty (by 7%) over a much shorter period [10].



FIGURE₁: CHANGES IN POVERTY HEADCOUNT IN SELECTED COUNTRIES.



FIGURE₂: PERCENTAGE OF CONTRIBUTION OF SERVICES SECTOR TO GDP IN 2006.

The above figure₂ shows Percentage of contribution of services sector to GDP in 2006. The highest contribution service from Tourism(Trade, Hotels and restarunt). In case of financial and business approx 5-6%. Financial service approximately 3-4%. In this paper we focus on providing funds to the low income households and reduce their poverty and increase their saving potential by Rotating Savings and Credit Association method. By introducing this method percentage of contribution of financial service sector to GDP will increase. This method will be useful to achieve the target of the MKUKUTA (reduce poverty between 1990 and 2015 by 50 %).

3 THE RANDOM AND BIDDING ROSCA SYSTEM

It is difficult to predict a random ROSCA as a series of loans and debt repayments [3]. The bidding allotment mechanism allows member to obtain a upatu when an unexpected opportunity or emergency arise, albeit at the cost of a discount [14]. Thus, in a certain world, a random Rosca is preferred by identical individuals desiring to overcome indivisibilities in consumption, while a bidding Rosca is superior in responding to heterogeneity among its members [13]. In contrast, the bidding ROSCA illustrated in Table₃ suggest that bidding ROSCA elements of lending and borrowing. There are 12 people who come together and form a group. Each one will contribute Tsh.10000/- per month and this will continue for next 12 months (equal to number of people in the group). In this group there will be one organizer, who will take the pain of fixing the meetings, collecting money from each other and then doing other procedures. To generate the pot of Tsh. 120000/- for example the organizer would gather six other participants and require that each (plus himself/herself) contribute Tsh.10000/- monthly. So each month all these 12 people will meet on a particular day and deposit Tsh.1,0000/- each. That will make a total of Tsh.120,000/- every month. At the end of the first rotation of contribution by seven participants, the organizer receives the Tsh.120000/- pot (including Tsh.10000/- of his own funds). In the second rotation naturally there will be few people who are in need of big amount because of some reason like some big expenses, liquidity crunch, business problem etc. Out of all the people who are in need of money, someone will bid the amount, depending on how desperate he/she is for this money. The person who submits the highest bid for the rotation receives the pot. Suppose out of total 6 people who bid for Tsh.500/-, Tsh.750/- and Tsh.1,000/-, the one who bids the highest will win. In this case it's the person (P2) who has bid Tsh.1,000/-. The high bid amount Tsh.1000/- will be deducted for the rest of the person (P3 to P12) so each person (P3 to P12) will contribute Tsh.9000/- for the second rotation. Hence participant P2 receives the total contribution for the second rotation Tsh.110000/-. So here you can see that the main winner took a big loss because of his desperate need of getting the money and others benefitted by it. So each person actually paid just Tsh.9000/-, not Tsh.10000/- in this case (they got Tsh.1000/- back). Note that when a person takes the money after bidding, he/she can't bid from next time, only rest of the people will be eligible for bidding. Now next month the same thing happens and suppose the best bid was Tsh.1500/-, then winner will get Tsh.106500/- this is illustrated in table₃ and the Tsh.1500/- will be reduced to people (P4 to P12). So each person (P4 to P12) is paying effectively Tsh.8500/-. This way each month all the people contribute the money, someone takes the money by bidding highest and the rest money is deducted back to members.

TABLE III
ROSCA BIDDING SYSTEM BASED ON TWELVE PARTICIPANTS

Members	Contribution by Month (in Tsh)												Total paid	Net gain (loss)
	High Bid	1000	1500	2000	1000	1500	1000	2000	1500	1000	2000	-		
Round	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12		
P1 (Organizer)	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	120000	0
P2	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	120000	(10000)
P3	10000	9000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	119000	(12500)
P4	10000	9000	8500	10000	10000	10000	10000	10000	10000	10000	10000	10000	117500	(13500)
P5	10000	9000	8500	8000	10000	10000	10000	10000	10000	10000	10000	10000	115500	(2500)
P6	10000	9000	8500	8000	9000	10000	10000	10000	10000	10000	10000	10000	114500	(3500)
P7	10000	9000	8500	8000	9000	8500	10000	10000	10000	10000	10000	10000	113000	2000
P8	10000	9000	8500	8000	9000	8500	9000	10000	10000	10000	10000	10000	112000	0
P9	10000	9000	8500	8000	9000	8500	9000	8000	10000	10000	10000	10000	110000	5500
P10	10000	9000	8500	8000	9000	8500	9000	8000	8500	10000	10000	10000	108500	9500
P11	10000	9000	8500	8000	9000	8500	9000	8000	8500	9000	10000	10000	107500	10500
P12	10000	9000	8500	8000	9000	8500	9000	8000	8500	9000	8000	10000	105500	14500
TOTAL RECEIVED	120000	110000	106500	104000	113000	111000	115000	112000	115500	118000	118000	120000	1363000	-

Note: Shaded cell represent the period in which the participants received the pot.

You will realize that the person who takes the money at the end will get all the money because there is no one else to bid now. So the person will get around Tsh.1,20000/- in the end, if you try to find out the returns which he/she got out of the whole deal, it will depend on things, how much higher bids were each month. If bids and charges are very low, then a person will make more money at the cost of other situations.

TABLE IV
CASH FLOWS OF A MEMBER
WHO TAKES THE LOAN IN THE SECOND MONTH

Month	Prized Amount/Loan In Tsh.	Contribution In Tsh
1	0	-10000/-
2	110,000/-	-10000/-
3	0	-10000/-
4	0	-10000/-
5	0	-10000/-
6	0	-10000/-
7	0	-10000/-
8	0	-10000/-
9	0	-10000/-
10	0	-10000/-
11	0	-10000/-
12	0	-10000/-

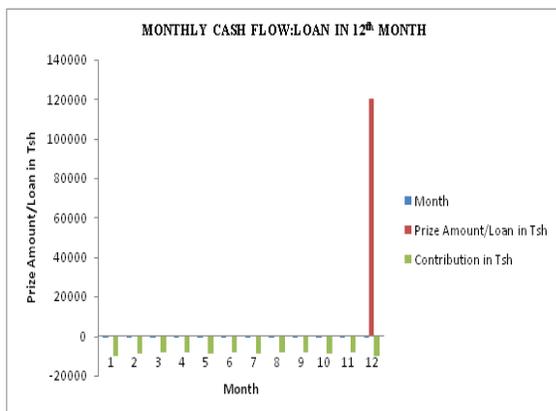


Figure4: Monthly Cash Flow in Second Month

Using the equation for net present value, the usual approach to finding a loan interest is to set the amount of loan received is equal to the sum of discounted present values of the monthly loan payments made for the term of the loan.

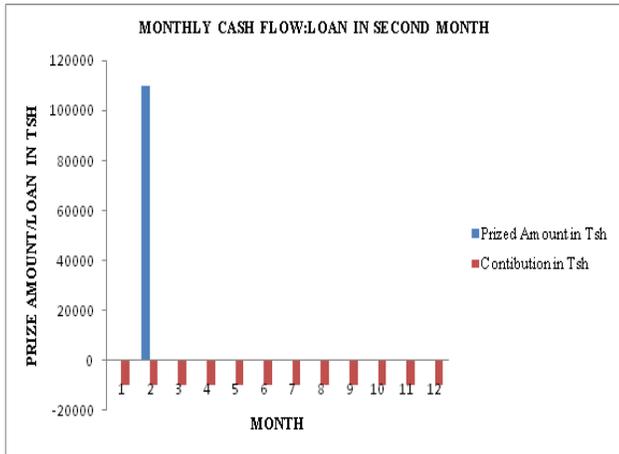


FIGURE 3: MONTHLY CASH FLOW IN SECOND MONTH

TABLE V

CASH FLOWS OF A MEMBER WHO TAKES THE LOAN IN THE SEVENTH MONTH

Month	Prized Amount/Loan In Tsh.	Contribution In Tsh
1	0	-10000/-
2	0	-90000/-
3	0	-8500/-
4	0	-8000/-
5	0	-9000/-
6	0	-8500/-
7	0	-9000/-
8	0	-8000/-
9	0	-8500/-
10	0	-9000/-
11	0	-8000/-
12	120,000/-	-10000/-

Using the below equation we can find the monthly interest rate is denoted by r

$$p = A \frac{1 - \left(\frac{1}{1+r}\right)^n}{r}$$

Where,

p - Principal amount borrowed

A - Periodic payment

r - The periodic interest rate divided by 100 (annual interest rate also divided by 12 in case of monthly installments.

n - The total number of payments (for a 30 years loan with monthly payments n=3*12=360.

4 ACKNOWLEDGMENT

We have taken efforts in this journal. However, it would not have been possible without the kind support and help of many individuals and organizations. We would like to extend my sincere thanks to all of them. We are highly indebted to Directors, principal, and member of St. Joseph College of Engineering and Technology, Dar Es Salaam, Tanzania for their guidance and constant supervision as well as for providing necessary information regarding the Journal and also for their support in completing the Journal. We would like to express my gratitude towards my parents and people of Tanzania for their kind co-operation and encouragement which help me in completion of this Journal. I would like to express my special gratitude and thanks to industry persons for giving me such attention and time. Our thanks and appreciations also go to my colleague in developing the Journal and people who have willingly helped me out with their abilities.

5 CONCLUSIONS

The results from this study underscore the importance of savings and borrowing vehicle for the poor and lower income households in Tanzania. ROSCA method is a very good tool for financing poor people. Especially in a ROSCA, members have quicker access to a large sum of money or a commodity than when saving individually. A further strength of the ROSCA is that it binds the members to save regularly, and thus to accumulate savings. An upatu is not a scalable model unless the upatu manager or company has sufficient personal resources as a backup for financial contingencies. The result in this paper provides strong support about benefit of saving and borrowing money for participants. In future work we plan to implement impact of setting up registered upatu in rural areas, impact of altering collateral/guarantee requirements, developing a credit scoring model for ROSCA. The day when the government and the industry participants alike understand the importance of ROSCA to the economy would mark the beginning of a new era.

REFERENCES

- [1] Rutherford, Stuart, and others. 1999. "Savings and the Poor: the Methods, Use and Impact of Savings by the Poor of East Africa." Report prepared for MicroSave-Africa, May 1999.
- [2] Calomiris, C. W. and I. Rajaraman, "The Role of ROSCAs: Lumpy Durables or Event Insurance?" *Journal of Development Economics* (1998), pp. 207-216.

- [3] Christy Chung Hevener Federal Reserve Bank of Philadelphia. November 2006. Alternative Financial Vehicles: Rotating Savings and Credit Associations (ROSCAs).
- [4] Stuart Rutherford Institute for Development Policy and Management University of Manchester. January 1999. The Poor and Their Money An essay about financial services for poor people.
- [5] Mary Kay Gugerty, Daniel J. Evans School of Public Affairs. "You Can't Save Alone: Commitment in Rotating Savings and Credit" Associations in Kenya, January 2005.
- [6] Ardener, Shirley. 1964. "Comparative Study of Rotating Credit Associations." Reprinted in Ardener and Burman, 1995.
- [7] Kimuyu, Peter Kiko. 1999. "Rotating Savings and Credit Organizations in Rural East Africa," *World Development*. 27(7): 1299-1308.
- [8] Chamlee-Wright, Emily, 2002. "Savings and Accumulation Strategies of Urban Market Women in Harare, Zimbabwe." *Economic Development and Cultural Change*.
- [9] Levenson, Alec and Timothy Besley. 1996. "The Anatomy of an Informal Financial Market: Rosca Participation in Taiwan." *Journal of Development Economics*, Volume 51.
- [10] Policy Forum "Growth in Tanzania: Is it Reducing Poverty?" www.policyforum-tz.org.
- [11] United Republic of Tanzania, 2002 Population and Housing Census.
- [12] Tanzania Human Development Report *The State of Progress in Human Resource Development* 1999. February 2000.
- [13] Besley, T., Coate, S., Loury, G., 1993. The economics of rotating savings and credit associations. *American Economic Review* 83, 792-810.
- [14] Stefan Klöpper "Understanding Chit Funds: Price Determination and the Role of Auction Formats in Rotating Savings and Credit Associations" *New Haven, Connecticut 06520-8269 USA*.
- [15] Bouman, F. J. A., 1979. The Rosca: financial technology of an informal savings and credit institution in developing economies, *Savings and Development* 4 (3), 253-276.

Designing an Automated Wheel chair with Stair Crossing Facility

Mst. Nasima Bagum, Choudhury Abul Anam Rashed, Sanjoy Kar

Abstract—now a day's all comforts in human life is the result of revolutionary advancement in technology. Modern Automation techniques; a part of technology is utilized in this design of "Automated Wheelchair" to give user a fully automated control. But the key story is not about control. The key approach was to find an easier way to cross inclined stairs with users. A new phenomenon was found from engineering analysis to cross stairs in an efficient way by the proper selection of wheel radius. The comfort is dedicated to the people who are suffering in normal movement. The developed 'Automatic Wheel Chair' would provide a comfortable and dynamic life-style to the handicaps which might be very much closer to their dreams.

Index Terms—Automation, Ergonomics, Handicaps, Power, Stairs, Wheel Chair, Wheel radius.

1 INTRODUCTION

A special type of wheel chair is designed which can be used to ride on stairs. So many works were done on automated wheelchair. But this factor was not considered or solved before in the way which would be discussed later. Emphasis was given on solving the problem as well as other facilities provided before. Giving ease to user in different postures is another important purpose [1]. The design presented in this paper will provide more physical and psychological advantages to the user because of scientific approaches and creativities. Our design will provide complete automatic control and higher mechanical efficiency. The most important thing is that, we included new scientific criteria (found by our analysis and experiment) that would provide the facility to ride on stairs that was not provided before. Obstacles in free movement are a curse to a human life. The wheel chairs available in the market provide them little facilities but it needs a lot of manual work done by users or helper to be operated. Its' movement area is limited also. This limitation creates a mental stress on a handicap specially those who are handicaps by accident [2]. These phenomena added fuel to fire in this thinking and this is why a highly efficient wheel chair is presented.

2 STATE OF THE ART

Dugas [3], invented the "Safer Automatic Wheelchair Wheel Locks". As the medical director of the Marion Nursing Home for some three decades, he became aware of the problem of wheelchair-related falls among semi-ambulatory patients who did not or could not remember to use manual brakes on their wheelchairs. Some patients in wheelchairs, particularly the elderly, have a tendency to fall and injure themselves when trying to stand because they fail to engage the manual locks. Dugas [3] hence attempted to find a way to save these patients from potentially debilitating injuries. He began his experimentation with locking systems and procured his first patent on April 20, 1993.

A motorized wheelchair, power-chair, electric wheelchair or electric-powered wheelchair (EPW) is a wheelchair that is propelled by means of an electric motor rather than manual power. Motorized wheelchairs are useful for those unable to

propel a manual wheelchair or who may need to use a wheelchair for distances or over terrain which would be fatiguing in a manual wheelchair. They may also be used not just by people with 'traditional' mobility impairments, but also by people with cardiovascular and fatigue based conditions [4].

Iturrate et al. [5] developed Non-Invasive Brain-Actuated Wheelchair Based on a P300 Neuro-physiological Protocol and Automated Navigation. Their research describes a new non-invasive brain-actuated wheelchair that relies on a P300 neuro-physiological protocol and automated navigation. When in operation, the user faces a screen displaying a real-time virtual reconstruction of the scenario and concentrates on the location of the space to reach. A visual stimulation process elicits the neurological phenomenon, and the electroencephalogram (EEG) signal processing detects the target location. This location is transferred to the autonomous navigation system that drives the wheelchair to the desired location while avoiding collisions with obstacles in the environment detected by the laser scanner. This concept gives the user the flexibility to use the device in unknown and evolving scenarios. The prototype was validated with five healthy participants in three consecutive steps: screening (an analysis of three different groups of visual interface designs), virtual-environment driving, and driving sessions with the wheelchair. On the basis of the results, this paper reports the following evaluation studies: 1) a technical evaluation of the device and all functionalities; 2) a users' behavior study; and 3) a variability study. The overall result was that all the participants were able to successfully operate the device with relative ease, thus showing a great adaptation as well as a high robustness and low variability of the system.

IBOT wheelchair Stair-climbing wheelchair was founded by inventor and entrepreneur, Dean Kamen [6] to transform the way people work and live. Dean Kamen's inventions always start the same way by looking at a problem, ignoring the conventional thinking that surrounds it, and working tirelessly

until it is solved. Before there was a Segway HT, Kamen and the researchers at his company DEKA developed the iBOT, the balancing wheelchair. The iBOT's code name was "Fred" or "Fred Upstairs" for the ability of the balancing and stair-climbing wheelchair to give the user the agility of the famous dancer, Fred Astaire. Like the Segway HT, the iBOT contains patented dynamic stabilization (iBALANCE) technology, an integrated combination of sensor and software components and multiple computers that work in conjunction with gyroscopes. Gyroscopes are motion sensors that help maintain balance. When the gyroscopes sense movement, a signal is sent to the computers. The computers process the information and tell the motors how to move the wheels to maintain stability. This electronic balance system is custom-programmed to the user's center of gravity, to monitor and respond to subtle changes in motion. Reach forward to shake hands, and the iBOT moves with you. Lean back and it moves away as well. The iBOT constantly realigns and adjusts its wheel position and seat orientation to keep the user upright and stable at all times, even when driving up and down curbs or inclines. In addition, the iBOT includes built-in triple redundant backup systems, as well as auditory and visual signals to provide even more safety and assurance. With input from the rider or an assistant, in "Stair Function" the iBOT utilizes gyroscopes and adjusts to the driver's center of gravity, climbing stairs by rotating wheels up and over each other. The iBOT can allow riders to stand up to the same eye-level as colleagues. The "Balance Function" of the iBOT can raise the rider to eye level for any number of business or social interactions. It lets the rider see over counters, and reach a high shelf in the office, kitchen or supermarket, safely and easily.

3 HUMAN FACTORS AND ERGONOMICS

We have to do analysis about the position of centre gravity of human body. Some experiments were made to find out the actual position of Center of gravity of human body. But from different experiments it was proved that the position shifts depending on postures. Even our prime concern of locating the position of "COG" in sitting posture on a chair varies due changes in following variables:-

- Sitting posture (Upright or Relaxed)
- Slope of inclined plane.
- Kinematic Status (Static or Dynamic)

S. Ashahara [7] from Japan did a significant job of determining the mean position of combined center of gravity of human and chair with standard deviations. For this purpose he divided his test areas of human body in 13 segments as shown in fig. 1. His experiment was good enough for determining the position in static conditions with a great accuracy. So he suggested a zone where the Combined "COG" would be found is about 2.5 inch to 6 inch from the axis passed through shaft of rear wheel perpendicularly [7].

In present calculation the "COG" of chair and human was considered separately because the chair's "COG" is fixed when the following factors are fixed:- Material and design as shown in fig. 2.

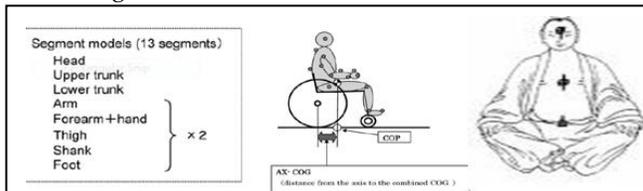


Fig. 1. Center of Gravity of human body

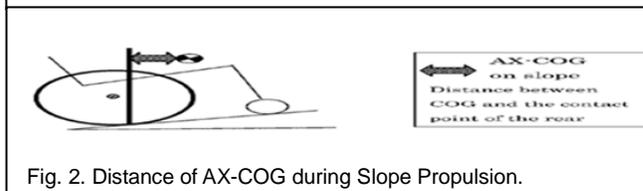


Fig. 2. Distance of AX-COG during Slope Propulsion.

Some other data were taken which are necessary to ensure comfort as shown in table 1.

TABLE 1
ANTHROPOMETRIC DATA

FACTOR	95 th	50 th
Functional forward reach	91.3	85.5
Buttock knee depth	67.2	64.2
Hip breadth	43.6	39.4
Elbow to elbow breadth	53.6	47.6
Sitting eye height	32.4	27.3
Popliteal height	51.8	47.2
Buttock popliteal depth	57.8	52.5

4 RESULTS AND ANALYSIS

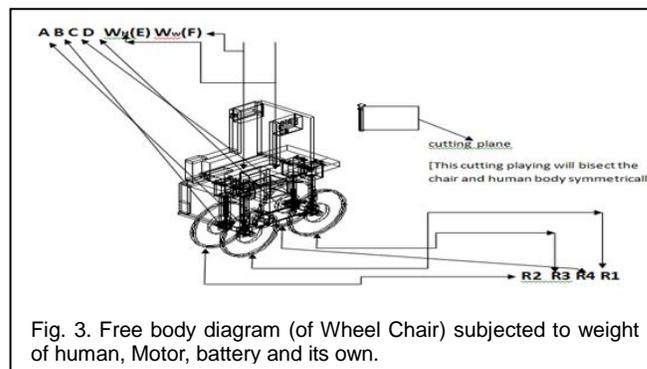


Fig. 3. Free body diagram (of Wheel Chair) subjected to weight of human, Motor, battery and its own.

Now; $\sum F_y=0$; so; $R_1+R_2+R_3+R_4= W_H + W_w$ [$R' = R_1+R_2$ and $R'' = R_3+R_4$.]

Here R' and R'' act at the midpoint of AD and BC respectively (P & q). Because of the symmetry of human body and the chair. $W_H = 154 \times 9.8 =$ Weight of human body. $1000 \times 9.8 = W_w =$ Weight of wheel chair, motor, battery. E= Center of gravity of human body; F= center of gravity of wheel chair (as shown in fig 3). W_H is taken based on mean weight of handicaps calculated from a sample of 12 handicaps and injured people. W_w is cal-

culated from the materials and design of the chair.

$$So; R' + R'' = W_H + W_w = 11310 \text{ N} \quad (1)$$

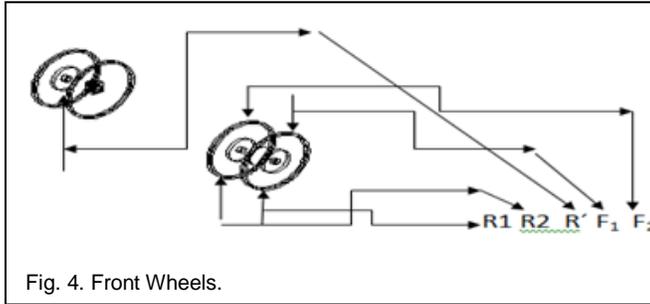
$$\sum M_p = 0; 0.36 \times W_w + 0.6 \times W_H + 0.9144 R' = 0$$

$$So, R'' = 4966.8 \text{ N} \quad (2)$$

Solving (1) & (2)

$R' = 6343.12 \text{ N}$ and $R'' = 4966.8 \text{ N}$ and $So; R_1 = R_2 = 2483.4 \text{ N}$ & $R_3 = R_4 = 3171.56 \text{ N}$

Front wheels



$$\sum F_y = 0; So, F_1 + F_2 = 4966.8$$

$$\sum M_A = 0; So AB(F_1 - R_1) = 0; So, F_1 = 2483.4 \text{ N}; So; F_2 = 2483.4 \text{ N}$$

Wheel connecting Rod

Shearing stress

$$\tau = \text{Force/Area} = F/2\pi L = .024 \text{ Mpa. Here; } r = .0381 \text{ M; } L = .43 \text{ M}$$

$$\text{Normal stress } \sigma = \frac{\text{Force}}{\pi r^2} = 0.187 \text{ MPa}$$

Ball-Bearing [At Position k and L]

$$[(\text{Sphere shape}) \text{ for } F_2 \text{ and } F_1] \text{ so; stress} = \frac{F_1}{4\pi r^2} = 0.725 \text{ Mpa;}$$

$$\text{stress} = \frac{F_2}{4\pi r^2} = 0.725 \text{ MPa}$$

Supporting plate (on ball bearing)

Normal stress, For $F_1, \sigma = F_1 / A = 0.57 \text{ MPa}$, for $F_2, \sigma = 0.57 \text{ MPa}$

Shearing stress, (for F_1 and F_2) = 3.42 MPa ; Deformation, $\delta_n =$

$$\frac{\delta L}{L} = 1.08 \times 10^{-6} \text{ m, } \delta s = \frac{\delta L}{L} = 1.08 \times 10^{-6} \text{ m, } E = 40 \text{ Gpa}$$

Force and stress on springs

Left leg and right leg Force = $F_1 = 3307.82$ [Actually $F_1 - (F_B + F_p + F_{i(B)})$ (Forces increase while jumping on stairs)]

$$\text{stress developed on springs } \tau = \frac{16PR}{\pi d^3} \left(\frac{4m-1}{4m-4} + \frac{.615}{m} \right)$$

$$= 366.42 \text{ Mpa}$$

Elongation

$$S = \frac{64pR^3 n}{Gd^4}$$

$$= 3.98 \times 10^{-16} \text{ m}$$

Right leg

$$\tau = 366.42 \text{ Mpa ; } S = 3.98910^{-10} \text{ m}$$

Load and stress on hollow rectangular support

Left/Right leg

F_1 or F_2

$$R = 1.25 \text{ in}$$

$$D = 5 \text{ in}$$

$$m = 2R/d = 4$$

$$G = 83 \text{ Gpa} = \text{modulus of Elasticity}$$

$$= 83 \times 10^9 \text{ NM}^{-2}$$

$$N = 10$$

Left Leg and Right Leg

$$\sigma = F / A = 0.157 \text{ MPa, } \tau = F/A = 0.09 \text{ MPa, } \delta_s = 5.17 \times 10^{-7} \text{ m}$$

$$\delta_n = 3.14 \times 10^{-7} \text{ m; [Here, } \sum F_y = 0, \text{ so, } So, F_4 + F_3 = R_4 + R_3.] \quad (3)$$

$$\sum M_c = 0, So, CD (F_4 - R_4) = 0 \quad (4)$$

From Eqn (3) and (4)

$$F_4 = 3171.56 \text{ N} = F_3$$

Backward/Rear Wheels-Wheel Connecting rod (stress),

$$\tau = 2.253 \text{ MPa, } \sigma = 0.099 \text{ MPa}$$

Ball Bearing for (F_3) and (F_4)

Circular Disc, $\tau = 0.8 \text{ MPa}$, Sphere, $\tau = 0.563325 \text{ MPa}$

Wheel (3) and (4)-

Shearing Stress $\tau = 0.145 \text{ MPa; } \sigma = 0.06993 \text{ MPa}$

Supporting Plate (F_3 and F_4) Backward legs

$$\sigma = 0.4428 \text{ MPa; } \tau = 2.657 \text{ MPa; } F_3 = F_4 = 3171.56 \text{ N}$$

Stress and deformation on spring $\tau = 284.708 \text{ MPa, } \delta = 3.09 \times 10^{-16}$

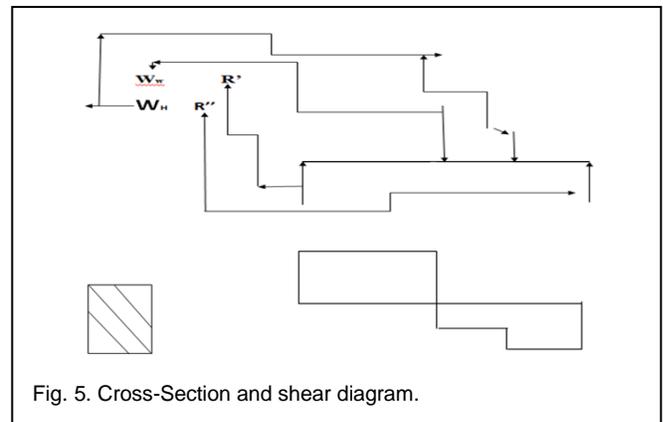
Load on hollow rectangular support (Left and right)

$$F = 2481.84 \text{ N; } \sigma = 0.121 \text{ MPa; } \tau = 0.069 \text{ MPa; } \delta_s = 3.95 \times 10^{-7} \text{ m; } \delta_n = 2.42 \times 10^{-7} \text{ m}$$

Flexure Stress on Seat

The shear fig 5 showing that maximum Bending moment hence Flexure stress occurs at point of action of Chair's weight.

$$\text{Max } M = 4966.8 \times 0.36 = 1787.77 \text{ N-M; Max } \sigma = \text{Max } M / S = 6^* \text{ Max } M / bh^2 = (6^* 1787.77) / (0.76^* 0.13^2) = 0.88 \text{ Mpa}$$



Dynamics (General and aerodynamics)

$$\sum F_y = 0; So; W_w + W_w = R_1 + R_2 + R_3 + R_4$$

For Movement; $\sum F_x \geq 0$

$$So, F_s \geq \mu_k (W_H + W_w) + 0.5 (C_d A Q V^2) \geq 3393.24 \text{ N}$$

$$P_{\text{wheel}} \geq F_s \times v \geq 3393.24 \text{ W}$$

[Here, $F_s =$ Force Supplied, $F_k =$ kinetic frictional force; $0.5 C_d A Q v^2 =$ Aero-dynamic resistance; $C_d = 0.50$ drag coefficient; $A = 0.8 \text{ m}^2$; $Q = 1.2$ (air); $v = 1 \text{ m/s}$, $\mu_k = 0.3$]

Moving through inclined plane as in fig 6 (For movement)

$$F_s \geq \mu_k (W_w + W_h) \cos \theta + (W_w + W_h) \sin \theta + F_{ar} \geq 11336 \text{ N}$$

$$P_s = F_s \times v \geq 7.59 \text{ HP [Where; } v = 0.5 \text{ ms}^{-1}; \theta = 30^\circ; \mu_k = 0.58]$$

Where; $v = 0.5 \text{ ms}^{-1}$ $\theta = 45^\circ$; $\mu_k = 1$

$P_s \geq 10.72 \text{ HP; So increasing the angle of inclination results in more power requirement.}$

Movement through Inclined Plane (Stairs) as in fig 7

Here a new criteria is developed about the force required to cross an obstacle of a certain height.

$$F_o \times CB > (w - F_s \sin \theta) \times CD$$

$$F_s \cos \theta \times CB > (w - F_s \sin \theta) \times \sqrt{2rh - h^2}$$

$$F_s > (w - F_s \sin \theta) \times \frac{\sqrt{2rh - h^2}}{(r - h) \cos \theta}$$

$$\text{Hence } O_f = \frac{\sqrt{2rh - h^2}}{(r - h) \cos \theta} = \text{Obstacle factor}$$

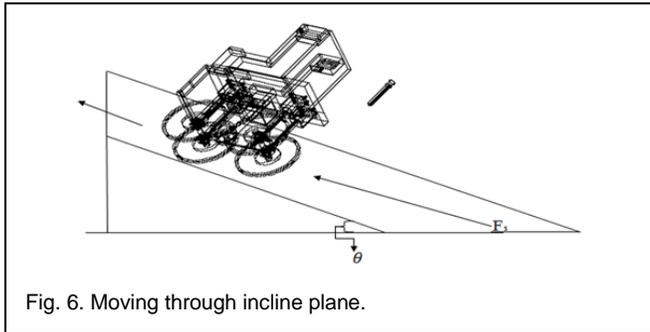


Fig. 6. Moving through incline plane.

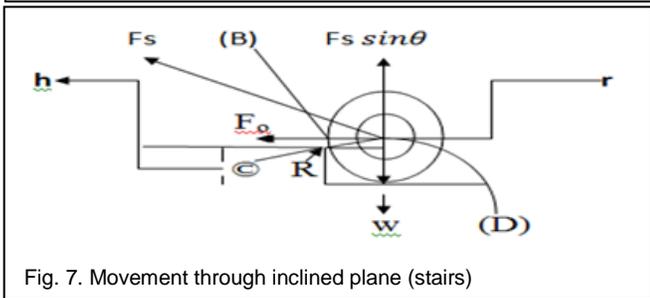


Fig. 7. Movement through inclined plane (stairs)

Here; r=radius of wheel; h= height of obstacle; R=reaction on point of contact; W=weight acting on wheel; F_o=Component of force (F_s) required to cross an obstacle of a certain height. So, radius should be greater than height of obstacle.

Condition -I as in fig-8 (All wheels try to cross the obstacle)

$$F_s \geq \mu_k(w_w + w_h) \cos \theta + (w_w + w_h) \sin \theta + F_{ar} + F_{c.o} \geq \mu_k(w_w + w_h) \cos \theta + (w_w + w_h) \sin \theta + F_{ar} + O_f \cdot \frac{w_w + w_h}{O_f \sin \theta + \cos \theta} \geq 25367.95 \text{ N}$$

So; P_s ≥ 10.20 HP; Where, v = 0.3ms⁻¹; θ = 30°; μ_k = 0.58; h = 5 inch; r = 7.5in

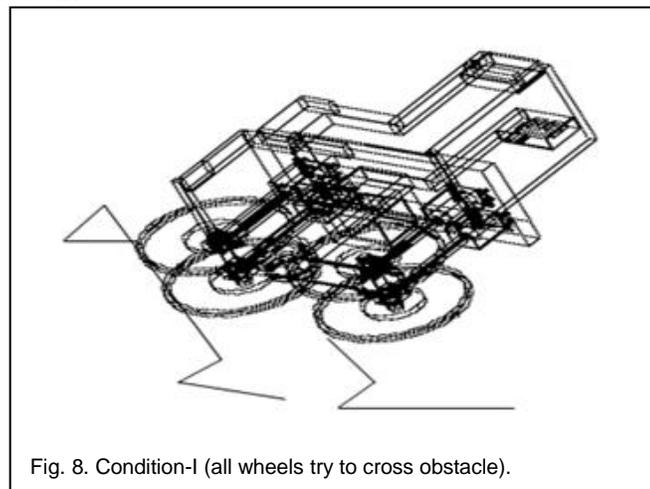


Fig. 8. Condition-I (all wheels try to cross obstacle).

Condition of minimum Force hence power supply when weight, angle of inclination and velocity is fixed. i.e. (dO_f / dr) = 0

That indicates the absence of Obstacle would result in lower

force requirement. But user can't ignore the presence of obstacle. So the design is focusing to an optimized approach of "increasing the wheel radius hence to increase the difference between radius and height of obstacle". But this approach would increase the total weight acting in the system. But here it can be proved the increment in radius would not result in a significant change in force requirement but significantly reduce the total force needed.

$$\frac{dF_s}{dr} = \frac{\left[\left\{ (r-h) \cos \theta + (\sqrt{2rh-h^2}) \sin \theta \right\} W \cdot \frac{2h}{\sqrt{2rh-h^2}} \right] - \left[\left\{ W \sqrt{2rh-h^2} \right\} \left\{ \cos \theta + \frac{2h \sin \theta}{\sqrt{2rh-h^2}} \right\} \right]}{\left\{ (r-h) \cos \theta + (\sqrt{2rh-h^2}) \sin \theta \right\}^2}$$

For θ = 30°, r = 7.5 in, h = 5 in, W = 11310N, dF_s / dr = -1065.618 N/in

$$\frac{dF_s}{dw} = \mu_k \cos \theta + \sin \theta + F_{ar} + O_f \cdot \frac{1}{(O_f \sin \theta + \cos \theta)}$$

For θ = 30°, μ_k = 0.58

$$dF_s / dw = 2.24$$

But increment of weight is 8.5 N/inch radius increased, so it results in 19.04 N/inch

So, dF_s / dr >> dF_s / dw

So, increase the radius within the size restriction. Here power is calculated in terms of velocity. But moment or Torque is more significant than the velocity because sufficient moment should be generated to overcome an obstacle against the moment due to weight.

Condition of Maximum moment (All wheels try to cross the obstacle)

$$\tau_s > W \sqrt{2rh - h^2}$$

Condition-II as like fig 9 (Front wheels try to cross the obstacle; Rear wheels moving on flat plane)

For r = 7.5 in, h = 5in, W = 11310N; τ_s > 2031.334N-M

F_s ≥ μ_k(R₁ + R₂) + F_{ar} + F_{c.o} ≥ 15546.144 N. So P_s ≥ 10.42 HP

For θ = 30°, r = 7.5 in, h = 5 in, R₁ + R₂ = 4996.8 N; μ_k = 0.58, V = 0.5.

$$F_{c.o} = (R_1 + R_2) \times (\sqrt{2rh - h^2}) / (r - h)$$

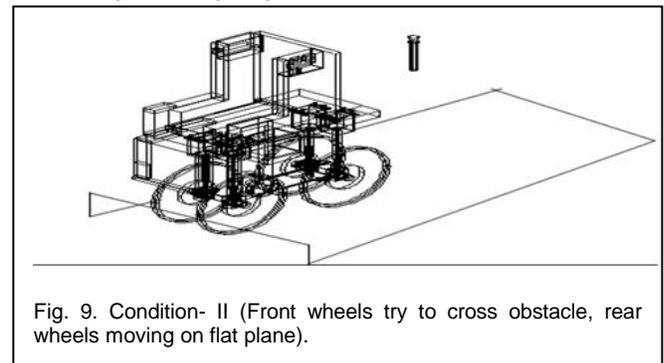


Fig. 9. Condition- II (Front wheels try to cross obstacle, rear wheels moving on flat plane).

Condition-III as like fig 10 (Rear wheels try to cross the obstacle; Front wheels moving on flat plane)

F_s ≥ μ_k(R₃ + R₄) + F_{ar} + F_{c.o} ≥ 11544.375N; So, P_s ≥ 7.74HP;

For θ = 30°, r = 7.5 in, h = 5 in, R₁ + R₂ = 4996.8 N; μ_k = 0.58, V = 0.5

$$F_{c.o} = O_f \cdot \frac{R_3 + R_4}{(O_f \sin \theta + \cos \theta)}$$

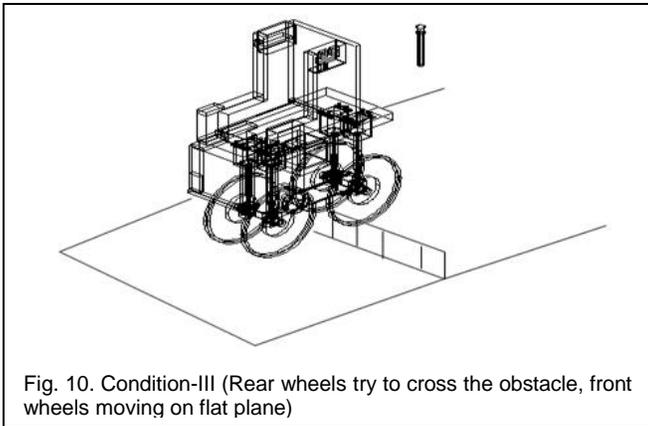


Fig. 10. Condition-III (Rear wheels try to cross the obstacle, front wheels moving on flat plane)

TABLE 2
STATISTICAL DATA FOR ANGLE OF INCLINATION, STAIR HEIGHT AND WEIGHT

No	θ	No	θ
1	20	7	32
2	22.8	8	33.8
3	23.4	9	36.47
4	25.1	10	41
5	27.2	11	43.87
6	28	12	45
No	$h(\text{inch})$	No	$h(\text{inch})$
1	4	7	5.4
2	4.3	8	5.5
3	4.48	9	5.7
4	4.5	10	6.2
5	4.55	11	6.48
6	4.6	12	7
No	$w(\text{kg})$	No	$w(\text{kg})$
1	78	7	101
2	87	8	113
3	88.9	9	132
4	89.3	10	134.6
5	93	11	138
6	95.5	12	154

For 50th Percentile, $\theta = 30^\circ$, $h = 5\text{in}$, $w = 98.25\text{ kg}$; For 95th percentile, $\theta = 45^\circ$, $h = 7\text{in}$, $w = 154\text{ kg}$ [9].

5 SOURCE OF POWER AND POWER TRANSMISSION

From the calculation it is found that the pulley connected to wheel should provide power greater than 10.72 HP. A 10kw (Cont.) is utilized here; 273.6 VDC Brushless DC servo motor / torque motor; 273.6 VDC as a power source. Torque motors are frameless kit motors. They consist of a permanent magnet rotor and a laminated stator. Cooling system is provided with Motor [10].

Motor Torque (Cont.) = 1200 N-M

Motor Outer Dia=19.69 in and Motor Length= 6 in (App.) This is the torque input in the smaller 3 inch Pulley. The power is transmitted to a 9 inch larger pulley by a V-Belt which is connected to the rear shaft.

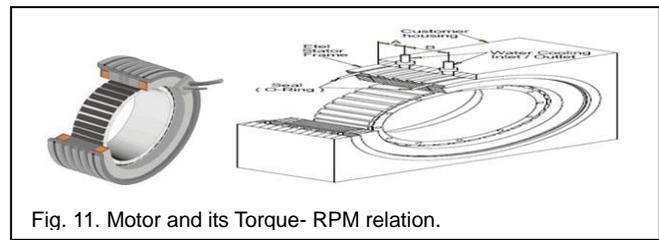


Fig. 11. Motor and its Torque- RPM relation.

$P_{in}=10000\text{ w}$; $\tau_{in}=1200\text{ N-M}$, $P_{out}=10000\text{ w}$ (App.) (slip and slack is negligible)

So, $N_{in} = 79.58\text{ rpm}$, $\tau_{out}=3600\text{ N-M}$; $S_0, N_{out}= 26.5\text{ rpm}$

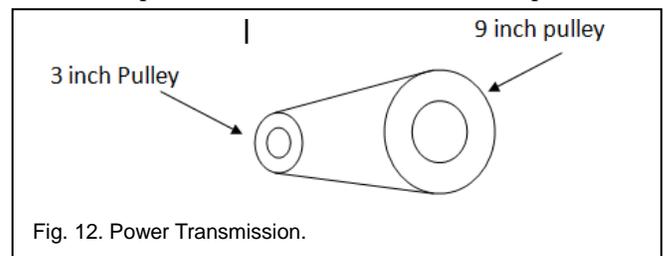


Fig. 12. Power Transmission.

$$\text{So, velocity of rear wheel } V = \frac{2\pi N_{out} * r}{60}$$

$$= 0.53\text{ m/s} > 0.3\text{m/s(req.)}, \tau_{out}= 3600\text{ N-M} > 2031.334\text{ N-M.}$$

These allowances are given due to prevent unknown (without experiment) reverse torque generated due to collision among wheels and stairs.

Voltage Applied to the DC poles; $V_{\text{applied}} = 273.6\text{ V}$, $R_{\text{armature}}=1.10(\text{Cu})$;

$$I = \frac{V_{\text{applied}} - V_{\text{cemf}}}{R_{\text{armature}}}$$

$$P = I \cdot V_{\text{cemf}} = P_{out} = 10000\text{W}$$
; So, $I = 44.517\text{ A}$

$P_{input} = V_{\text{applied}} * I = 12180\text{ w}$ (App.) ; **Heat generated** = $I^2 * R = 2180\text{ w}$ (App.) ; **Efficiency** = $10000/12180 = 82\%$ (App.); **Torque of L.P**

= $(T_1 - T_2) * r_1 = 3600\text{ N-M}$; **Torque of S.P** = $r_2 (T_1 - T_2) = 1200\text{ N-M}$

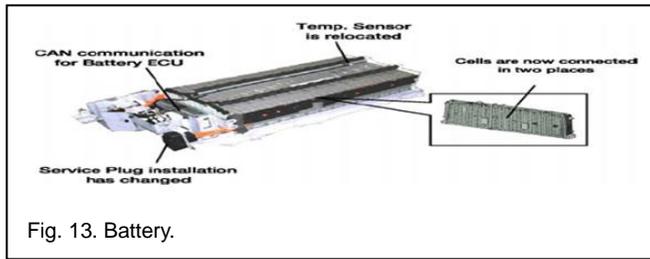
Here $(T_1 - T_2) = 31500\text{ N}$ (App.) is the resultant force in the belt.

So belt should be selected based on:-

- Cross sectional area of V-belt
- Max tolerable stress
- Stress developed, Torque, RPM
- Length

Now A **Nickel Metal Hybrid Battery** is used as a voltage hence power source to motor [11].

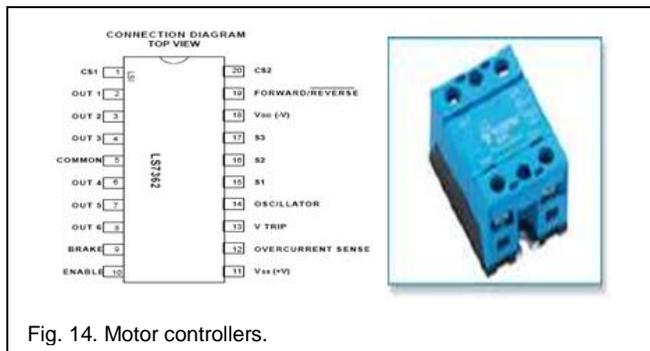
The HV battery pack contains six nickel-metal hydride 1.2V cells that are connected in series to form one module In the 01-03 Prius, 38 modules are divided into two holders and connected in series. Thus the HV battery contains a total of 228 cells and has a rated volted of 273.6V. The electrode plates in the HV battery are made of porous nickel and metal hydride alloy.



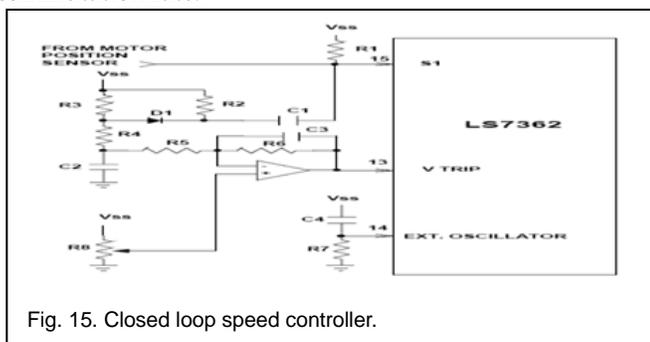
Both motor and battery manufacturer provide cooling system. In the present design the facility of containing and circulating cooling fluid (Liquid and air) is provided.

Motor Controller (Features)

- Speed Control by Pulse Width Modulating (PWM) only the low-side drivers reduces switching losses in level converter circuitry for high voltage motors.
- Open or closed loop motor speed control.
- Externally selectable input to output code for 60° 120° 240°, or 300° electrical sensor spacing.
- Three or four phase operation.
- Analog Speed control.
- Forward/Reverse control.
- Output Enable control.
- Positive Static Braking.
- Overcurrent Sensing.
- Six outputs drive switching bridge directly.

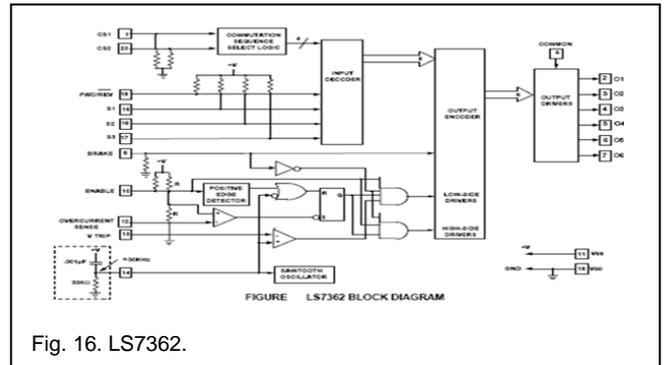


The COMMON, Pin 5, is tied to the positive supply rail and LS7362 Output 1, 2, and 3 are used to drive level converters Q101, Q102 and Q103, respectively. Only the motor top side drivers consisting of Q107, Q108 and Q109 Which are connected to the motor supply, VM, Will be subject to the high speed switching currents that flow through the motor. The level converters are turned on and off at the slower commutation rate.



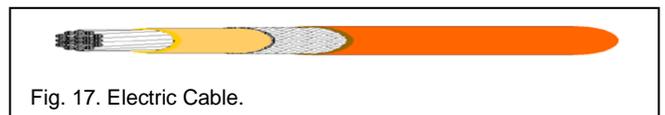
Input/output Description: Commutation Selects (Pins 1, 20)

These inputs are used to select the proper sequence of outputs based on the electrical separation of the motor position sensors. With both inputs low (logic zero), the sequence is adjusted for 60° electrical separation with CS2 high and CS1 low 120° separation sequence is selected, with CS1 high and CS2 low 240° separation sequence is selected and with CS1 and CS2 high the 300° separation sequence is selected. Note that in all cases the external output drivers are disabled for invalid SENSE input codes. Internal pull down resistors are provided at Pins 1 and 20 causing a logic zero when these pins are left open.



Recommendation for Electric Cables

EXRAD FX 600 volt shielded battery cable designed specifically to handle the higher voltage and current. The jacket insulation isolates any stray currents making this cable very safe. Thin wall and high temperature insulations allow for lower weight and less space. EXRAD FX 600 volt shielded battery cable is able to withstand temperatures of 240°C and higher [13].



6 OTHER OPERATIONS

$$E_r = \text{Energy for rotation}; E_f = \text{Energy for friction}; S = r \cdot \alpha; \alpha = \pi/4; r = 0.216m. \text{ For moving the front wheel in an angle; } \alpha = 45 \text{ deg}$$

$$E_r = 2 \times \frac{1}{2} I \omega^2 = m n^2 W^2 = m r^2 \left(\frac{v^2}{r^2}\right) = m n^2 \left(\frac{s^2}{t^2 n^2}\right)$$

$$= \frac{m s^2}{t^2} = \frac{m s^2}{1^2} = m s^2 = \left(\frac{R s^2}{g}\right) = 18.61W$$

$$E_f = \mu_k \cdot R \times S = 323.50W; E_t = E_f + E_r = 342.11W;$$

This would be supplied from battery by using **H-bridge** for both rotations. One 1 KW DC motor is used for this operation. Cylindrical Rods standing on Wheel's shaft are connected with a V-belt Drive Consists of identical pulley to maintain same speed of rotation of both wheels.

H-bridge is an electronic circuit which enables a voltage to be applied across a load in either direction. These circuits are often used in to allow DC motors to run forwards and back-

wards [14].

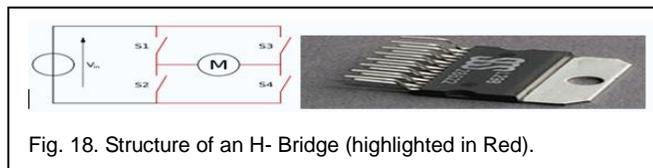


Fig. 18. Structure of an H- Bridge (highlighted in Red).

Seat-back angle control

Switching the carriage also include magnetism in magnet and pulls down the lock. Here another electromagnetic mechanism is used by switching control board. we can move the carriage back and fro to get preferred a seat back angle.

Power Needed $S = r\theta = 0.48m$

$$W = FS = 31.65J; P = 10.54W; [T = 3S; F = 65.90N(\mu R)]$$

Seat length control

This is not an automatic control. Before starting to use the user will set the length. The back support is on a rectangular carriage which can be moved through a passage containing locking system to set the length in desired positions.

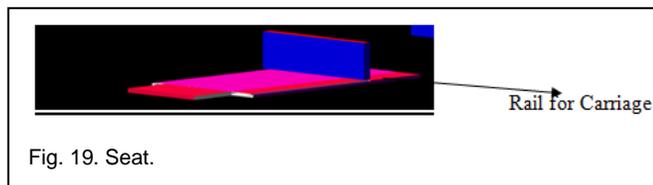


Fig. 19. Seat.

7 MECHANICAL BRAKES

For safety a mechanical brake is used .It would help the motor brake to stop a heavy load (Big momentum) easily. Hydraulic System is used. It can be controlled by hand.

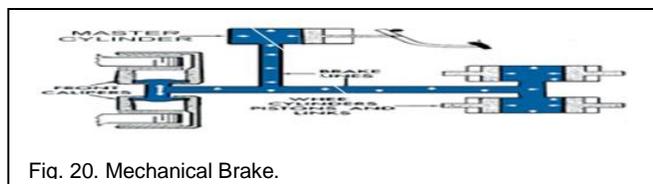


Fig. 20. Mechanical Brake.

8 TIRES

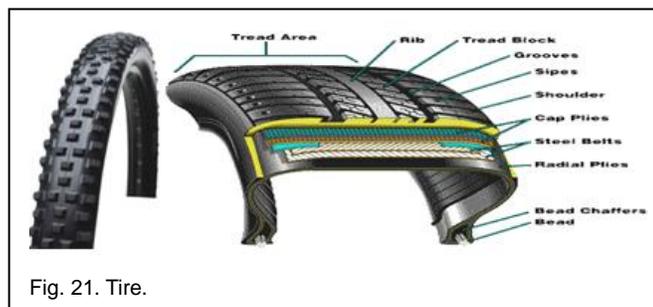


Fig. 21. Tire.

Tires are used to make frictional grip to resist slipping because some user may have tiles fitted in stairs. They are selected based on rating to provide necessary frictional grip hence Coefficient of friction as stated before [15].

9 MATERIALS

Low cost Grey cast Iron for Body and frame (providing Damping capacity and required load bearing capacity) [16], Respective materials for accessories as recommended, Cotton Foams for seat and Lead free painting and coatings are required.

10 COST ANALYSIS

The costing is for a single automated wheel chair.

TABLE 3
INDIVIDUAL PARTS COST

Brushless DC Motor (10 kw)	400\$
DC Motor(1 Kw ;)	50 \$
MFG. COST	200 \$
Battery (1 small,1 large)	2500\$ ~ 3000\$
Gray Cast Iron	450~550\$
Controllers, Cables and Circuits	150\$
Cooling and Lubricating Liquids	3\$/L
Magnet; Bearings	47\$
Paintings and Coatings	10\$
Tires	50\$
Belt Drive and Accessories	30\$
Mechanical Brake	150\$
Others	60\$
Total	4100\$~4700\$

Statements that serve as captions for the entire table do not need footnote letters. By reducing the power requirement (Battery, Motor etc.), Mass production and location of warehouse could be cost reducing factors efficient Sourcing could reduce cost

11 DISCUSSION

The wheel chair was developed by considering different principles of science and engineering and our innovative approach. But due to high cost we could not test it practically. This is why emphasis was given on developing the theoretical background by considering all known factors which can affect the functioning of the product.

12 CONCLUSION

In developed nations almost all buildings have the facilities of elevators and separate stair ways for wheels. But these facilities are rarely found in homes of South Asian, African Nations. A narrow band of customers are focused who can afford to buy this facility of leading a self dependent (In movement) life being a handicap. Further improvement can make this chair more outstanding and to overcome the remaining shortfalls.

REFERENCES

- [1] M. Halender, "A Guide to the Ergonomics of Manufacturing," 1st East-West Edition, Publisher, Taylor & Francis, pp. 21 ~ 23
- [2] <http://helpguide.org/mental/post-traumatic-stress-disorder-symptoms-treatment.htm>, 2004
- [3] G. A. Dugas, "Safer Automatic Wheelchair Wheel Locks," http://en.wikipedia.org/wiki/Motorized_wheelchair, 1993.
- [4] www.freepatentsonline.com/4898256.html , 2008
- [5] I. Iturrate, J. Antelis, A. K'ubler and J. Minguez, "Non-Invasive Brain-Actuated Wheelchair Based on a P300 Neurophysiological Protocol and Automated Navigation," <http://webdiis.unizar.es/~jminguez/>, 2007
- [6] D. Keman, "IBOT wheelchair Stair-climbing wheelchair," <https://www.msu.edu/~luckie/segway/iBOT/iBOT.html>, 2000.
- [7] S. Ashahara, "Research work on Combined centre of Gravity of human and wheelchair," unpublished
- [8] F.L. Singer, "Strength of Materials," 4th Edition, Harper & Row, Publishers, Inc. , pp. 131~139, 1987
- [9] <http://en.wikipedia.org/wiki/Percentile>, 2009
- [10] G. A. McCoy, "Energy-Efficient Electric Motor Selection Handbook;Direct Drive Torque Motors for Machine Tool Applications", pp. 217-219, 2004
- [11] "Toyota Hybrid System" course-071, section-3, pp. 3-1 to 3-8
- [12] <http://en.wikipedia.org/wiki/LSIComputerSystem-LS7362>, 2010
- [13] www.champcable.com/pdf/EXRAD-Hybrid , 2009
- [14] http://en.wikipedia.org/wiki/H_bridge , 2008
- [15] www.bridgestone.co.in/tyre/tyreknowledge/index.asp , 2010
- [16] B.K. Agarwal, "Introduction To Engineering Materials," Tata McGraw Hill, New Delhi, India, pp. 209~216, 1997

A Case Study on Implementation of Prince2 Methodology in Automotive Industry in Malaysia (A Preliminary Study)

Saima Saad, Abdullah Ibrahim, Omme Asma, Muhammad Saad Khan and Junaid Akhtar

Abstract - Operations & Management division is the backbone of any corporate organization and it is the key factor to improve the performance of the industry. Global competition of automotive industry required well-organized project management in each area for satisfaction of customer and quality. The new approach of Project Management is Project Management PRINCE2 (PMP2), is the generic framework, which is design to suit every type of project in industry. Therefore, the aim of this research is to investigate the impact of project management prince2 methodology in Malaysian automotive industry. PRINCE2 is a structured method for effective project management. PRINCE2 has established in UK as generic, process based approach to project management. In view of the fact that its emphasis on, the dividing the project into manageable and controllable stages. This research will encourage the implementation of Prince2 methodology in automotive industry of Malaysia. Additionally, it will encourage the prospect utilizations of PMP2, which ultimately improve the human performance in automotive industry.

Index Term -- automotive industry, Prince2 methodology, project management.

1 INTRODUCTION

Organizational project management is the systematic management of projects, programs, and portfolios in alignment with the accomplishment of strategic goals. The concept of organizational Project Management is based on the idea that there is a correlation between an organization's capabilities in Project Management, Program Management, and Portfolio Management, and the organization's in implementing effectiveness strategy.

As businesses change at a faster rate, it is becoming increasingly important to execute on projects. [1] Additionally, due to the broad nature of much of the change, projects are affecting larger parts of the organization. Therefore, just as the need to perform projects is increasing, the complexity in executing them is also increasing. Organizational Project Management draws from the broad base of project management and organizational design applications to understand the organizational processes that affect the ability to manage the delivery of projects.

Manufacturing companies are under increasingly diverse and mounting pressure due to more sophisticated markets, changing customer choice, and global competition. With globalization broadening the marketplace and increasing competition, customers are placing greater demands on manufacturers to increase quality and flexibility while maintaining or decreasing costs (Dangayach and Deshmukh 2003).

Manufacturing is no more concentrated in one country, but rather spread over distant locations across the globe. In such a competitive scenario, companies have to search for new processes, materials, suppliers/vendors, manufacturing facilities locations, and delivery channels for their products and services at a competitive price. The advanced/new manufacturing technologies have harnessed a wide range of benefits, including reduced costs, increased

productivity, greater flexibility, and higher quality, enabling companies to improve their competitive position. By adopting an appropriate manufacturing strategy, companies can achieve excellent manufacturing status and compete effectively in global markets.

Operations & Management division is the backbone of any corporate organization, more so in this era ever advancing technology. Management of project or task is very essential in daily routine work. Organizations have to organize and schedule their work by using the project / task management to achieve their task successfully.

The uncertainly associated with the project failure has forced organizations to adopt a structured and process based approach to project management. Bellis, (2003) suggested that the structured project management means managing the projects in a logical organized way, following defined steps. A structured project management method is the written description of this logical, organized approach.

Egginton, (1996) further argued that an effective process based approach can be defined as 'the one which brings together the most fundamental principles of project management in a way that overcomes differences and altogether maximizes the operational effectiveness of the organization, measured in terms of delivery to time, within the budget, to specification and within maximum customer satisfaction'.

Researchers, like Bellis, (2003) argue that the structured process based approach contributes to a better understanding of the ultimate goal and output of the projects, provides a common language to all the stakeholders involved in the project, provides flexible decision points, generic framework and methodology to reduce the uncertainties associated with the projects and encourages formal recognition of responsibilities within the project.

2 AUTOMOTIVE INDUSTRY IN MALAYSIA

The automotive industry in Malaysia has developed since the establishment of Proton in 1985, followed by Perodua in 1993 as a part of the National Car Project. The introduction of the National Car Project has given a boost to the development of components and parts manufacturing in Malaysia. Currently, there are four local vehicle manufacturers including Proton, Perodua, Naza, and Modenas. In addition, there are nine motor vehicle assemblers and 343 components-parts manufacturers in Malaysia [12]. Despite fluctuation in automotive production, the vehicle production in Malaysia tends to increase due to the rapid increase in domestic sales. The total vehicle production in 2007 is 441,678 vehicles as compared to 360,105 vehicles in 2000. From January to March 2008, the total number of vehicles production is 132,744.

The number of vehicle sold in the domestic market is 487,176 in 2007 as compared to 343,173 vehicles in the year 2000. In 2008, it is forecasted that the number of vehicles sales will grow by more than 5%. The total industry sales are predicted to achieve 580,000 vehicles in 2012. Most of the vehicle sales in Malaysia are dominated by the local manufacturers. Perodua has the highest market share at 33.3% with total sales of 162,152 vehicles in 2007. Proton has the second biggest market share in 2007 with total sales of 118,134 vehicles or 24.2% market share. This is followed by Toyota with 81,993 vehicles (16.8%), Honda 28,478 vehicles (5.8%), Naza 20,286 vehicles (4.2%), Nissan 18,569 vehicles (3.8%), Inokom 9874 vehicles (2.0%), followed by Daihatsu, Mitsubishi, Mercedes, Hyundai, BMW, Kia, Isuzu, Ford, and Suzuki (less than 1% each) [12].

3 THE PRINCE2 METHODOLOGY

Project management PRINCE2 is a structured method for effective project management. PRINCE2 has established in UK as generic, process based approach to project management. It is a de facto standard used extensively by UK government and it has started to be recognized and used in the private sector, in UK. PRINCE2 the method is in the public domain, offering non-proprietary best practice guidance on project management and it is starting to emerge internationally. [8]

Hall, (2003) suggest that PRINCE 2 provides an easily tailored and scalable method for the management of all types of projects. Uncertainty and change are the important factors that underpin the adoption of PRINCE2 methodology by the organizations. There are always many changes during the life of the project, people change their mind, and requirements change. These affect what the project is doing. PRINCE2 has a technique of controlling the way changes impact on the project in order to prevent the project going off in the wrong direction.

PRINCE2 is derived from an earlier method called PROMPTII and from PRINCE project management method, which was initially developed in 1989 by the Central Computer and Telecommunications Agency (CCTA) as a UK Government standard for information systems (IT) project management; however, it soon became regularly

applied outside the purely IT environment. PRINCE2 was released in 1996 as a generic project management method. PRINCE2 has become increasingly popular and is now a de facto standard for project management in the UK. Its use has spread beyond the UK, Holland, Denmark, Australia and other countries.

Since 2006, the method has been revised and launched as "PRINCE2:2009 Refresh" in 2009. The name "PRINCE2" (instead of "PRINCE3" or similar) is kept to indicate that the method remains faithful to its principles. Nevertheless, it is a fundamental revision of the method from 1996 to adapt it to the changed business environment, to make the method simpler and "lighter", to address current weaknesses or misunderstandings, and to better integrate it with other methods.

Following Case Studies of implementation of PMP2 for excellence in Human performance

- A project case study different from the other case studies which describe an organisation's use of PRINCE2 in that it is a project manager's description of the use of PRINCE2 on a specific project. The Registers of Scotland provided it. [9]
- The Cheshire Constabulary Case on 2002, Senior level commitment to PRINCE2 ~ Strong business focus of IT projects ~Implementing a project mentality Electricity Supply Board Ireland is the national electricity utility in the Republic of Ireland.[9]
- In April 2002 Business process re-engineering using PRINCE2 ~ Achieving business goals with PRINCE2 ~ Converting to the euro using PRINCE2
- Use of PRINCE2 processes, components and techniques on the Enterprise ~ Risk Management Project ~ A PID template ~ Sample Highlight Report including Resource Usage Summary, Project Deliverables list and Checkpoint Report Sample End Stage Report including Risk Log and Quality Log .
- The Getronics PRINCE2 - PMI/PMBOK Combination Case Study in 2003

PRINCE2 - PMI/PMBOK Combination Case Study based on material supplied by Getronics. Brief summary of the two approaches, Current perceptions of relative positioning. Getronics view of combining the two approaches and the complementary benefits. Max Wideman's detailed comparison

4 PROBLEM STATEMENT

Automotive companies all over the world are under increasingly diverse and mounting pressure due to markets that are more sophisticated, changing customer choice, and global competition. Global competition of automotive industry required well-organized project management in each area for satisfaction of customer and quality. Implementation of fresh project management process based approach "Prince2 methodology" for increasing of human performance in automotive industry.

5 JUSTIFICATION OF THE PROBLEM

The dynamic nature of the business environment has forced the organizations especially in process industries to experience extraordinary levels of change. Those organizations who fail to manage the inherent risk associated with change, innovation and management of projects often end up with high proportion of project failures. Change has become a way of life for organizations that need to remain effective and competitive in order to thrive (Rietveld, 2004). Those organizations who fail to manage the inherent risk associated with change, innovation and management of projects often end up with high proportion of projects failures.

The global automotive industry is a significant constituent of industrial and economic advancement, and its development has considered global competitiveness of leading industrialized economies. This industry is a reasonably developed one and involves huge investments in research and development and technology. It is also seen as an indicator of the economic progress of a country. [10, 11]

6 OBJECTIVES

- To analyse the existing project management problems associate with Malaysian automotive industry.
- To analyse the effectiveness of present project management system applicable in Malaysian Automotive industry.
- To implement and measure the value of PMP2 in automotive industry for achieving the excellence in human performance.

7 METHODOLOGY

This research will base on an extensive critical literature review of the project management principles, methodology and its practical implications in automotive industry. Methodology of prince2 consists of following important stages of the project, which are directing, starting, initiating, controlling, managing stage boundaries, closing and product delivery. However, most of the things are inter-related with each other as shown in following figure 1.

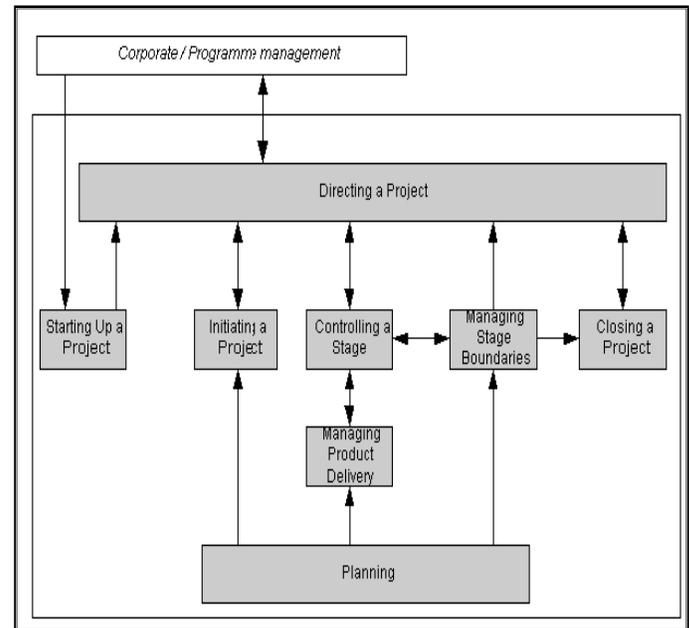


Figure1: Process model of PRINCE2 Project Management.

The theme of PRINCE2 methodology is the generic framework, which is designed to suit every type of project in any industry. PRINCE2 does not cover every aspect of project management tools and techniques. The techniques and tools will vary according to the project type and corporate environment.

The specialized aspects that are excluded from PRINCE2 are shown below:

- People management techniques such as motivation, delegation and team leadership.
- Generic planning techniques such as Gantt charts and critical path analysis.
- Risk management techniques.
- The creation and management of corporate quality management and quality assurance mechanisms.
- Business case management, budgetary control and earned value analysis.

The exclusion of specialized tools and techniques and a generic approach for every type of projects in automotive industry highlights important issues that needs to be investigated.

8 EXPECTED RESULTS

The research on prince2 methodology in automotive industry will provide a following expected result of the study

- The research will investigate and identify the problems associated with the automotive industry in Malaysia.
- Implementation of Prince2 methodology based on dividing the project into manageable and

controllable stages, which will help to increase the human performance in automotive industry.

- The research will promote the implementation of Prince2 methodology in the future.

9 ACKNOWLEDGMENT

Authors would like to pay sincere gratitude to University Malaysia Pahang for providing Financial support through the funding, Vote No. GRS110330has gratefully acknowledged.

10 REFERENCE

- [1]D. Milosevic and P. Patanakul, "Standardized project management may increasedevelopment projects success,"Int. J. of Proj.Managm vol. 23, pp. 181-192, July2004.
- [2] Hazell Peter et.al. , "Property rights, collective action and technologies for natural resources management". The Commons in an Age of Globalization, the Ninth Biennial Conference of the International Association for the Study of Common Property; 17-21 June 2002.
- [3] Morris Peter. "Managing the Front-End: how project managers shape business strategy and manage project definition". Global Congress Proceedings – Edinburgh, Scotland, 2005.
- [4] Day W.J., "Project Management and Control". pp 194, Macmillan, 1994.
- [5] Egginton B.; "Multi-National consterium based projects improving the process". Int. journal of Project management, Volume 33, pp 58-64, 2002.
- [6] Hall D.; "Implementing the prince2 Methodology."2003.
- [7] David H et.al. "A project contingency framework based on uncertainty and its consequences." International Journal of Project Management, vol 28, pp 256–264, 2010.
- [8] www.wikipedia.com.
- [9] <http://www.projex.com/prince2-case-studies>, October 2011.
- [10] T. Laosirihongthong et.al. "A Comparative Study of Implementation of Manufacturing Strategies in Thai and Indian Automotive Manufacturing Companies." Journal of Manufacturing Systems Vol. 24, No. 2, 2005.
- [11] G.S. Dangayach. S.G. Deshmukh, "Evidence of manufacturing strategies in Indian industry: a survey." Int. J. Production Economics vol 83, pp 279–298, 2003.

[12] Malaysian Automotive Association. Summary of sales and production data.Malaysia Automotive Association, October 2011.
http://www.maa.org.my/info_summary.htm

[13] Rosli M and Kari F."Malaysia's national automotive policy and the performance ofproton's foreign and local vendors". Asia Pacific Business Rev.; vol 14, no 1, pp 103–18, 2008.

[14] Ministry of International Trade and Industry Malaysia. Opening of the Malaysian Automotive Association New office, Ara Damansara. Ministry of International Trade and Industry Malaysia.
http://portaluat.miti.gov.my/cms/content.jsp?id¼com.tms.cms.article.Article_7cffba1a-c0a81573-11c011c0-34fd0b7c, 2008.

Abortion and Challenges of Teenage Pregnancy in Lagos, Nigeria.

By

Atere, Adewole Akinyemi
Department of Sociology and Criminology
Osun State University
Okuku Campus, Osun State

Ayodele, Johnson Oluwole
Department of Sociology,
Faculty of Social Sciences,
Lagos State University,
Ojo – Lagos, Nigeria.

and

Omololu, O
University of Ibadan,
Faculty of the Social Sciences
Department of Sociology
Ibadan, Oyo State

Abstract

The study examined the influence of abortion on the lives of pregnant teenage girls in Lagos. The study adopted an exploratory and descriptive research design. Copies of a questionnaire were administered on one hundred and five girls selected from Ojo Local Government Area of Lagos State who participated in the study. The study conducted ten in – depth interviews involving medical experts, parents and teenagers who were purposively selected. These data were analysed. Findings of the study supported the outcomes of previous studies that adolescent girls are hesitant to discuss their past exposure to abortion; have divergent perceptions of abortion and display inadequate though varying knowledge of the implications of abortion for their future reproductive health. Moreover, the study found evidences of cultural and religious inhibitions that discourage girls from public discussion of their sexual behaviour. Scared by the disgrace that such an act might cause, most girls deliberately do not access abortion services from specialists. Therefore, the study suggests that government legalises abortion to enable experts handle abortion and post abortion needs of adolescent girls. This will salvage many innocent babies and girls who might be involved in unsafe abortion from experiencing fatality. Otherwise, quackery in the health sector may continue to undermine all progressive health initiative to lay siege on the lives of vulnerable unborn babies and the life chances of their adolescent mothers in Lagos.

Key Words: Adolescent Mothers, Life Chances, Post-Abortion Services, Unborn Babies, Unsafe Abortion.

1. Introduction

The World Health Organization (WHO) has defined adolescence as progression from the onset of secondary sex characteristics to sexual and reproductive maturity; development of adult mental processes and adult identity and transition from socio-economic dependence to relative independence (WHO, 1975 cited in Population Reports, 1995:3). One third (36.5 million) of Nigeria's total population of 123 million are youth between the ages of 10 and 24 (Population Reference Bureau, 2000). Though the age range of adolescence is culture and purpose specific, in Indonesia many studies on adolescent reproductive health have defined adolescents as young people aged 15-24 years (Yayasan & Kantor, 1993; Utomo, 1997; LD-FEUI, 1999; Situmorang, 2001).

Approximately one billion people—nearly one out of every six persons on the planet — are adolescents; 85 percent live in developing countries (UNFPA. 1997). The sexual vitality of most adolescents puts them at risk of various reproductive health challenges. While about 15 million adolescents aged 15–19 years give birth each year, as many as 4 million obtain an abortion (UNAIDS, 1997). Lack of sexual health information and services for these vulnerable citizens places the young people at risk for pregnancy and unsafe abortion having implications for life and living. If over 16 percent of teenage females reported first sexual intercourse by age 15 (National Population Commission, 2000); a study of 330 female rape victims in Benin

City, Nigeria, reported that a majority of rape victims were females ages 13 to 19; 48 percent were under age 13 and in another study, 75 percent of rape victims were revealed to be unmarried and ages 13 to 19 (Omorodion & Olusanya, 1998), it is urgent that their pregnancy terminating behaviour especially through illegitimate mechanisms deserve some scientific examination and analysis.

Abortion practices remain continuously controversial in many parts of the developing countries of the world because it is intricately linked up with morality than adolescents' liberty to live their lives the way they consider desirable. Many people, especially religious leaders have argued that providing reproductive health information and services to adolescents will encourage them to become promiscuous. As a result, the approach of successive governments to the question of adolescent reproductive health tends to be based on morality, rather than the health need. It is probably for this reason that performing or seeking an abortion has been and still remains illegal in Nigeria, except it is intended to save a woman's life. Irrespective of public policy attachment to moral sentiments, experts estimate that more than 600,000 Nigerian women obtain abortions each year (Henshaw et al., 1998). One study found that one-third of women obtaining abortions were adolescents. Though students also demonstrated an increase in knowledge of contraceptive options (Centre for Communication Programs, 1995), hospital-based studies showed that up to 80 percent of Nigerian patients with abortion-related complications were adolescents (Otoide et al., 2001).

Adolescent unwanted pregnancies often end in abortion. Surveys in developing countries show that up to 60 percent of pregnancies to women below age 20 are mistimed or unwanted (International Council on Management of Population Programmes, 1997). Pregnant students in many developing countries often seek abortions to avoid being expelled from school (Zabin & Kiragu, 1998). Induced abortion often represents a greater risk for adolescents than for older women. In Nigeria, for example, 50–70 percent of women hospitalized for complications of induced abortion are younger than 20; a 13-year review found that complications from unsafe abortion were responsible for 72 percent of maternal deaths among women under age 19 at one university hospital (Unuigbo et al., 1988).

Apart from the sensitivity of the issues, studies in Jakarta found that most parents felt inadequate to talk to their children about issues related to reproductive health (Iskandar, 1995; Utomo, 1997). Parents do not know how to deal with their children's sexuality any better than young people know how to deal with it themselves. Today young people are increasingly tolerant of premarital sex. A study among young people aged 15-24 from various socio-economic backgrounds in 12 cities in Indonesia in 1993, revealed that between 10 to 31 percent of youth reported having engaged in premarital sex (YKB, 1993). Young people, especially those unmarried, seldom use contraception. Sexually active single young people who have sex with a steady partner often claim that intercourse is not the result of premeditated or conscious decisions but just "happens", so they are unlikely to be prepared with contraception (Khisbiyah et al., 1997). In addition, many young people have limited knowledge of contraception (LD-FEUI, 1999).

In Indonesia, as abortion is restricted and childbearing out of wedlock is unacceptable, many premarital pregnancies result in marriage. However, when marriage is not an option, many girls turn to abortion. Since abortion is illegal, it is often performed by unskilled providers in unsafe conditions (Adrina et al., 1998:126; Indraswari, 1999; 131-164). It is estimated that up to half of all pregnancy – related deaths in Indonesia result from the complications of unsafe abortion (Muluk, 1994 cited in Mohamad, 1998:84). Also, a study carried out by the Centre for Health Research at the University of Indonesia in 2000, estimated around 2 million abortion cases per year in Indonesia and roughly 30 percent of them were for adolescents (Utomo, et al., 2001), the situation may not be significantly different in Nigeria.

Unsafe abortions that take place each year are 20 million, mostly in countries where abortion is illegal. World health organisation and Guttmacher Institute (2002) claim abortion is safe in countries where it's legal, but dangerous in countries where it's outlawed and performed clandestinely. According to WHO, nearly all abortions (92 per cent) are safe in developed countries, whereas in developing countries, more than half (55 per cent) are unsafe. As a result, the World Health Organisation recognises unsafe abortion as a

“silent pandemic” and believes safe and legal abortion is a fundamental right of women, irrespective of where they live (WHO & Guttmacher Institute, 2007). Regrettably however, for Nigerian women, abortion is a luxury. For instance, Chapter 21 of Nigerian Criminal Code ‘Offences against Morality’ criminalises abortion. *Section 228* of The Nigerian Criminal Code provides that any person who, with intent to procure miscarriage of a woman whether she is or is not with child, unlawfully administers to her or causes her to take any poison or other noxious thing, or uses any force of any kind, or uses any other means whatever, is guilty of a felony, and is liable to imprisonment for fourteen years.

Abortion is not a typically Nigerian challenge; it is

2. Material and Methods

The study location was Lagos. The survey covered all the state’s statutory 20 Local Government Areas. In every local government area, 5 respondents were selected, giving a total of 100 respondents. The study area was chosen for its level of urbanization and diverse characteristics. Lagos derives its demographic importance and prominence from being a premier city having considerable political and economic functions. It has a population of about (9.2m) which accommodates over 6.2 percent of the national population of 140 million (2006 population census figure). At 9 percent per annum growth rate, approximately 300,000 persons per annum or 25,000 per month or 34 persons per hour are added to the existing population (Noah, 2000). Metropolitan Lagos is the most heterogeneous city in the country. It remains the economic nerve centre of the country and industrialized city in the country.

Both qualitative and quantitative methods were used for data gathering. For quantitative data the Questionnaire method was used and for qualitative, In-depth interview was used for the collection of data. In order to make the sample size representative of the whole population in the study area, random sampling technique was used.

3. Data Collection

This study is based on a mixture of qualitative and quantitative approaches. In addition to empirical data collected through in – depth interviews in Lagos, available materials were reviewed for related information. The main data collection techniques for primary data gathering employed in this study are: survey and

nonetheless, a dominant social affliction especially in developing societies of the world. A qualitative study in Yogyakarta in 1997 among 44 women who had a premarital pregnancy at age 15-24 had 18 respondents terminating the pregnancy (Khisbiyah et al., 1997:43). Of those who continued the pregnancy, 21 respondents married during their pregnancy and only five respondents remained single. Facing the fact of their pregnancy, most girls who decided to continue their pregnancy had attempted abortion (usually by drinking traditional medicine/jamu) but failed. PKBI reports indicate that for 1998 and 1999, in Indonesia, there were two million abortions each year, and 750,000 (38 percent) of them were requested by single young women (Media Indonesia, 2000).

in – depth interview. One hundred girls were randomly selected from Ojo Local Government Area in Lagos State to participate in the study. For the study, three categories of respondents were selected for five in – depth interviews: two medical experts, two parents (one male and one female) and one teenager who were purposively selected

Five graduating students of Sociology Department of the Lagos State University, comprising three ladies and two men were engaged as assistants who conducted the in – depth interviews and administered copies of the questionnaire with the researcher. These assistants were rigorously trained on the techniques of interviewing people in ways that will prevent respondents from holding any vital information back. Pre interview practice sessions were held to avail the research assistants the competence to administer questionnaire, handle tape recorders and transcribe their contents. These orientation exercises took five days. In all, one hundred copies of a questionnaire were administered, retrieved, correctly completed and analysed for the study.

For qualitative data, five in – depth interviews were conducted to elicit qualitative information about the challenges which the reproductive health of adolescent girls faces in Lagos state. Participants were selected across varying socio – economic backgrounds within the study area. The survey covers the extent of the influence of abortion on the life chances of pregnant adolescent girls in Lagos state. Consent forms were administered to all the participants before the interview. The researcher safely keeps tapes, consent forms and notes taken during the in – depth interview.

4. Data Analysis

Quantitative Data

Returned questionnaires were subjected to thorough

editing. Prior pre-coding of the questionnaire facilitated entry and analysis. Quantitative data collected from this survey were subjected to analysis. This involves an examination of the distribution of the respondents according to particular characteristics. This decision is informed by the assumption that behaviour of individuals in society is, to a large extent, determined by their personal characteristics as well as those of the environment in which they live their daily lives and their abortion behaviour.

5. Results

Socio – Demographic Variables

From table 1, the age pattern of the respondents indicated the proportion of respondents in age group 20 years below is proportionally more than other categories. This implies that majority (48%) of the sample population are in the sexually. While 63% of respondents are single, 25% are married, 8% are divorced and 4% separated. With 48% of respondents having the first school leaving certificate and 44% holding the West African school certificate and 8% holding Ordinary National Diploma, illiteracy could possibly not be advanced as an acceptable premise for adolescent pregnancy and unsafe abortion embracing behaviour of girls in Lagos state. Among respondents, there are more Christians than Muslims. While 58.5% are Christians, 35.5% are Muslims, 3% are traditional believers and 4% belongs to others. The level of ethnicity did not deviate significantly from the expected pattern as 55% are Yoruba, 25.5% Ibos; 9% Hausas and others 11%.

Teenage Pregnancy and Abortion Behaviour

Table 2 reveals that 8% of respondents believed that girls began to experience sexual activity before they were 10years old. While 34% said that girls began between age 10 and 14 years, 58% maintained that it was between age 15 and 18 years. One significant inference from these is that parental influence on girl child morality has manifestly ebbed. Criminalisation of abortion is not a recent legal initiative. The traditional attitude of parents to adolescent pregnancy has always predisposed the pregnant girls to the patronage of all sorts of illegitimate abortion inducing devices.

An old woman respondent said,

The level of moral depravity in contemporary society

combines with children's superficial bodily structure to fire the thoughts of members of the opposite sex that result in most of the unsafe pregnancies and the abortion risks that follow.

While 83% of respondents believe that adolescent girls were aware of the legal implication of their actions, only 9% noted that they were unaware and 8% observed that they were indifferent. Adolescent girls have significant decision making power in terms of whether to live with their accidental pregnancies or induce abortion. Fifty nine percent of respondents said partners of the adolescents introduced them to abortion; 19% said friends; 43% said nobody and 13% mentioned parents. That 43% of respondents adjudged adolescent girls as being solely responsible for the decision to induce abortion indicate their decision making skill in what happens to their premature motherhood is not completely outside their control.

A parent respondent said,

Looked at from any angle, pregnant adolescents are decision makers even if they lack required experience that could make going to hospitals for abortion services a wise decision, abortion has two possible social results: it either acts as an eye opener or become a way of life.

Justification for Unsafe Abortion

Pregnant adolescent girls have a multiplicity of reasons for inducing abortion as 29% respondents adduced avoidance of stigma of being accidental mothers; 33% attributed it to avoidance of economic burden; 10% advanced freedom to work as the reason for inducing abortion; 17% ascribed it to the freedom to enjoy themselves; and 11% insisted that it was due to their desire to maintain their girlhood. No matter what the justification for inducing abortion is, *section 229* states that any woman who, with intent to procure her own miscarriage, whether she is or is not with child, unlawfully administers to herself any poison or other noxious thing, or uses any force of any kind, or uses any other means whatever, or permits any such thing or means to be administered or used to her, is guilty of a felony, and is liable to imprisonment for seven years. In addition, *section 230* of the code provides that any person who unlawfully supplies to or procures for any person anything whatever, knowing that it is intended

to be unlawfully used to procure the miscarriage of a woman, whether she is or is not with child, is guilty of a felony, and is liable to imprisonment for three years. Are pregnant adolescent girls ignorant of the existence of the above anti abortion law?

A male respondent said,

If an adolescent girl is impregnated by someone who is not her boyfriend and is a blood relation, the shame of the event might make the family aid the inducement of abortion because of its incestuous character.

What then could be responsible for this development among adolescents, Devine et al. (1993) found that parental divorce during early adolescence was associated with earlier onset and greater frequency of sexual activity for females, but not for males. Less parental supervision in single-parent homes and modelling of single-parent dating behaviour are among the possible explanations for these tendencies (Dornbusch et al., 1985; Miller & Moore, 1990).

Inherent Dangers of Abortion

Abortion that is induced in controlled environment by an expert could be a safe experience. But the situation alters when it is done in unhealthy condition and handled by quack doctors. Commenting on medical procedure for ensuring this, Dada (2011) noted that there are several ways by which abortion can be done, and there are different stages with their accompanying risks. If for some reason a medical abortion does not work, a woman has the option of getting it done surgically. On the desire to induce abortion, 25% of the respondents said some pregnant adolescent girls will prefer to die to inducing their abortions in the hospital; 35% concluded that out – of – hospital induction of abortion does not instantly lead to death; 30% believed that even adolescent girls who had their abortions induced at the hospital still sometimes die; 7% noted that a person will die when his/her time comes and 3% said pregnant adolescent girls have no reasons whatsoever for inducing their abortions outside orthodox hospitals.

On complications arising from unsafe abortions, a medical doctor respondent said:

Complications which could be either immediate or prolonged included severe blood loss leading to death; infections leading to infertility; uterus perforation;

cervical laceration; renal failure; damage to the bladder, intestine perforation and chronic pelvic inflammatory diseases.

Supporting the claim that abortion is a major cause of injury and death among women worldwide, another medical doctor respondent observed:

“In Nigeria, about 25 out of 1,000 pregnancies are aborted, compared to western countries where abortion is legal and done liberally. There, abortion rate is five out of 1,000 pregnancies. Non-legalisation of abortion contributes immensely to maternal deaths. We are losing about 60,000 women every year to abortion and 40 per cent of this figure is caused by complications from induced abortion.”

Another medical doctor respondent commenting on complications arising from unsafe abortions said, *Complications could be either immediate or prolonged. These included severe blood loss leading to death; infections leading to infertility; uterus perforation; cervical laceration; renal failure; damage to the bladder and intestine perforation. Prolonged effects of unsafe abortion include infertility and chronic pelvic inflammatory disease.*

Prevention and safeguards against teenage abortion

On safeguards against abortion, 43% of respondents recommended the provision of medical and counselling services by government to be the most auspicious panacea; 12% pointed to the direction of recreational facilities; 9% suggested preventive intervention; 3% favoured free education and 33% preferred the deepening of parental socialisation.

On the prevalence of abortion in Nigeria, an in – depth interview respondent said:

A poor family may not have any option than terminate the pregnancy of their teenage daughter who is in school especially when it is apparent that the means of livelihood for both pregnant adolescent and her adolescent ‘accidental husband’ might be precarious.

Consequently, thirteen percent opted for the need for parents to counsel their young girl children more appropriately; 15% suggested the use of professional counsellors to do the prevention; 35% raised their

thumbs up for improved family life education; 4% believed that improved girl child parenting at home would work wonders while 33% insisted that if parents pay more desirable attention to their growing girl children, the incidence of adolescent girl pregnancy and its tragic twin partner – unsafe induction of abortion will decline.

Another respondent advises co-parents and government,

While we should educate our pregnant adolescent girls, be more intimate with them and their unwanted babies so as to enable them overcome the impact of their traumatic conditions, government, should legalise abortion to prevent unnecessary death of young children.

How then can teenage pregnancies be prevented? Even in the West, the prevention of teen pregnancy has become a subject of great interest to policymakers in recent years. Educational programs delivered in the public schools, often as part of the overall health curriculum, have been among the widespread approaches to the problem (Kirby & Coyle, 1997).

6. Discussion

What option does a helpless pregnant girl have when on realising that she is pregnant; her parents drive her to meet the boy who impregnated her who is still wholly dependent on his own parents? In Nigeria, the fact remains that except abortion is committed to save the life of a pregnant woman; the Criminal Code legislates varying terms of imprisonment from 3-14 years for the person who masterminds an abortion, the woman and the person who supplies the instruments for the abortion. Criminalising abortion legally disempowers qualified doctors in Nigeria and leaves the fate of women with unwanted pregnancies in the hands of bizarre quack doctors.

With the rising susceptibility of adolescent girls in Lagos to accidental pregnancy and consequent unsafe abortion, there is a need for a second thought about legalising abortion. To date, almost all parents in the study area are intolerant of adolescent pregnancy and its attendant abortion. Parents are mostly usually in-

involved in abortion arrangement at the level of incest or the health of the pregnant girl is threatened.

While the effects of abortion are insurmountable ranging from physical, social, emotional to psychological effects, some adolescents might be very lucky to go through the process of abortion without complications (Dada, 2011). In spite of the manifest risks involved in the practice, pregnant adolescent girls still induce abortion. What could inform this risky lifestyle? The findings from some studies in the west indicate that children who live with both biological parents are less likely to be sexually active than those from one-parent homes (Flewelling and Bauman, 1990; Forste and Heaton, 1988; Hayes, 1987; Miller and Bingham, 1989; Newcomer and Udry, 1987; Upchurch et al., 1998; Whitbeck et al., 1996) while the sexual activity of children in remarried families usually falls between the first two (Miller & Moore, 1990; Thornton & Camburn, 1987).

7. Conclusion

Nigerian policy makers have no plausible justification for their indifference to youthful pregnancy that ultimately, in most cases, leads to wanton loss of life. Teenage pregnancies and consequent abortion will continue in so far as in some socio – cultural settings, sexual exposure of girls occurs at ages as young as 9 – 12 years as older men seek young girls as sexual partners to protect themselves from STD/HIV infections (Blanc & Way, 1998). To discontinue this unwholesome trend, public policy must urgently consider legalising abortion as a critical health issue rather than a moral question. Liberalising abortion is an issue that is basic to women's reproductive health. Sincerely, it transcends hypocritical morality.

8. Recommendation

It is recommended that government should evolve and implement an adolescent reproductive health policy that would save unborn children from the carnage that abortions subject them. Therefore, the suggestion to give reproductive health information and services to single young people may never be better at any other time than now. It is against this background that the need for improved health and social services aimed at adolescents, including reproductive health services, is not only being advertised by developed countries but should be increasingly recognized throughout the

world.

Policy makers should be committed to a conscious reduction by half, the conception and birth rates among under – 18s. The need for such measures is seldom questioned; in the public imagination (and often in the research literature), there is a belief that teenage pregnancy and childbearing are increasing and that action is necessary to stem the growing numbers of young mothers (FPSC, 1999).

9. Reference

- 1) Adrina, Kristi Purwandari, Nike Triwijati, and Sjarifah Sabaroedin. 1998. Hak-hak Reproduksi Perempuan yang Terpasung (Women's Reproductive Rights that are Repressed). Jakarta: Pustaka Sinar Harapan.
- 2) Center for Communication Programs. (1995). Reaching Young People Worldwide: Reproductive Health Communication Activities to Date, 1986-1995. Baltimore, MD: Johns Hopkins, 1995.
- 3) Dada, J. F. (2011). Abortion: any merit. Retrieved Friday August 26, 2011 from <http://www.tribune.com.ng/index.php/strictly-femail/22232-abortion-any-merit>
- 4) Devine, D., Long, P. and Forehand, R. (1993). "A prospective study of adolescent sexual activity: Description, correlates and predictors." *Advances in Behaviour Research and Therapy* 15:185-209.
- 5) Dornbusch, S. M., Carlsmith, J. M., Bushwall, S. J., Ritter, P. L., Leiderman, H., Hastorf, A. H. and Gross, R. T. (1985). "Single parents, extended households, and the control of adolescents." *Child Development* 56:326-341.
- 6) Flewelling, R. L., and Bauman, K. E. (1990). "Family structure as a predictor of initial substance use and sexual intercourse in early adolescence." *Journal of Marriage and the Family* 52:171-180.
- 7) Forste, R. T., and Heaton, T. B. (1988). "Initiation of sexual activity among female adolescents." *Youth and Society* 19:250-268.
- 8) FPSC (Family Policy Studies Centre) (1999). Teenage Pregnancy and the Family, Family Briefing Paper 9. London: FPSC.
- 9) Hayes, C. (ed.) (1987). *Risking the Future: Adolescent Sexuality, Pregnancy, and Childbearing*, Vol. I. Washington, DC: National Academy Press.
- 10) Henshaw, S. K. et al. (1998). The incidence of induced abortion in Nigeria. *International Family Planning Perspective* 1998; 24:156-164.
- 11) Indraswari. 1999. Fenomena kawin muda dan aborsi: gambaran kasus (The Phenomena of teenage marriage and abortion: a case study) in Syafiq Hasim (ed.) *Menakar 'Harga' Perempuan: eksplorasi lanjut atas Hak-hak Reproduksi Perempuan dalam Islam*. Bandung: Penerbit Mizan.
- 12) International Council on Management of Population Programmes (ICOMP). (1997). Adolescents/youth reproductive health hazards. *Feedback* 23(3):5
- 13) Iskandar, M. B. (1995). Laporan Akhir Pengembangan Module KIE Materi Kesehatan Reproduksi Untuk Keluarga dengan Anak Usia Sekolah dan Remaja (The Final Report of Development of Reproductive Health EIC Module for the Family with School Age Children and Adolescent), Jakarta: Pusat Penelitian Kesehatan UI, BKKBN and UNFPA.
- 14) Khisbiyah, Y., Desti, M., & Wijayanto. (1997). *Kehamilan tak Dikehendaki di Kalangan Remaja (Unwanted Pregnancy among Adolescents)*. Yogyakarta: Pusat Penelitian Kependudukan Universitas Gadjah Mada.
- 15) Kirby, D. and Coyle, K. (1997). "School-based programs to reduce sexual risk-taking behaviour." *Children and Youth Services Review* 19:415-436.
- 16) LD-FEUI. (1999). *Baseline Survey of Young Adult Reproductive Welfare in Indonesia 1998/1999. Executive Summary and Recommendations Program*. Jakarta: Demographic Institute Faculty of Economics, University of Indonesia (LD-FEUI).
- 17) Miller, B. C., and Moore, K. A. (1990). "Adolescent sexual behaviour, pregnancy, and parenting: Research through the 1980s." *Journal of Marriage and the Family* 52:1025-1044.
- 18) Miller, Brent C., and C. Raymond Bingham (1989). "Family configuration in relation to the sexual behavior of female adolescents." *Journal of Marriage and the Family* 51:499-506.
- 19) Muluk, (1994). in Mohamad, Kartono. 1998. *Kontradiksi dalam Kesehatan Reproduksi (Contradiction in Reproductive Health)*. Jakarta: Pustaka Sinar Harapan.

- 20) Newcomer, S. F, and Udry, J. R. (1987). "Parental marital status effects on adolescent sexual behaviour." *Journal of Marriage and the Family* 49:235-240.
- 21) Noah, A. O. (2000). *Fundamental of general studies*. Ibadan: Rex Charles publications.
- 22) Omorodion, F. I. & Olusanya, O. (1998). The social context of reported rape in Benin City, Nigeria. *African J Reproductive Health* 1998; 2(2):37-43.
- 23) Otoide, V. O. et al. (2001). Why Nigerian adolescents seek abortion rather than contraception: evidence from focus-group discussions. *International Family Planning Perspective* 2001; 27:77-81.
- 24) Situmorang, A. (2001). *Adolescent Reproductive Health and Premarital Sex in Medan*. PhD thesis, The Australian National University, Canberra.
- 25) Thornton, A. D., and Camburn, D. (1987). "The influence of the family on premarital sexual attitudes and behaviour." *Demography* 24:323- 340.
- 26) UNAIDS. (1997). *Report on the Global HIV/AIDS Epidemic: December 1997*. <http://www.unaids.org/highband/document/epidemioreport97.html> (accessed November 1998).
- 27) UNFPA. (1997). *UNFPA and Adolescents*. New York: UNFPA (1997). Also available at <http://www.unfpa.org/PUBLICAT/TECH/ADOLESES.HTM>.
- 28) Unuigbo, J.A. et al. (1988). Abortion-related mortality in Benin City, Nigeria: 1973 – 1985. *International Journal of Gynecology and Obstetrics* 26:435–439
- 29) Upchurch, D. M., Levy-Storms, L., Sucoff, C. A. and Aneshensel, C. S. (1998) "Gender and ethnic differences in the timing of first sexual intercourse." *Family Planning Perspectives* 30:121-126.
- 30) Utomo, I. D. (1997). *Sexual attitudes and behaviour of middle-class young people in Jakarta*, PhD thesis, The Australian National University, Canberra.
- 31) Whitbeck, L. B., Simons, R. L. and Goldberg, E. (1996). "Adolescent sexual intercourse." In Simons, R. L (ed.), *Understanding Differences Between Divorced and Intact Families: Stress, Interaction and Child Outcome: 144-156*. Thousand Oaks, CA: Sage.
- 32) Yayasan, K. B. & Kantor, M. N. K. (1993). *BKKBN/KLH. Hasil Need Assessment Reproduksi Sehat Remaja di 12 kota di Indonesia (Result of Assessment of Adolescents Reproductive Health Needs in 12 cities in Indonesia)*. Jakarta.
- 33) Zabin, L and Kiragu, K. 1998. Health consequences of adolescent sexuality and fertility behavior in sub-Saharan Africa. *Studies in Family Planning* 29(2):210–232.

RESULTS

Table 1: Socio - Demographic Characteristics of Respondents

Characteristics	Frequency	Percent
Sex		
Male	32	32
Female	68	68
Total	100	100
Age		
20 Below	42	42
21- 25	15	15
26 – 30	12	12
31 – 35	10	10
36 – 40	13	13
41 – and above	8	8
Total	100	100
Marital Status		
Single	63	63
Married	25	25
Widowed	8	8
Separated	4	4
Total	100	100
Education		
FSLC	48	48
WASC	44	44
OND	8	8
Total	100	100
Religion		
Christianity	58	58.5
Islam	35	35.5
Traditional	3	3.0
Others	4	4.0
Total	100	100
Ethnic Group		
Ibo	25	25.0
Hausa	9	9.0
Yoruba	55	55.0
Others	11	11.0
Total	100	100

Source: Author's Field Survey, 2011

Table 2: Abortion Behaviour of Pregnant Adolescent Girls in Lagos

Characteristics	Frequency	Percent
Age at which sexual activity begins		
Under Ten Year	8	8
10 – 14 years	34	34
15 – 18 years	58	58
Total	100	100
Awareness that inducing abortion outside the saving of lives was criminal		
Aware	83	83
Not aware	9	9
Indifferent	8	8
Total	100	100
Why teenagers induce abortion		
Avoidance of stigma	29	29
Avoidance of econ. burden	33	33
freedom to work	10	10
Freedom to enjoy themselves	17	17
Desire to maintain girlhood	11	11
Total	100	100
Who introduced adolescents to abortion		
partners	59	59
parents	13	13
friends	19	19
nobody	43	43
Total	100	100
Reasons for abortion outside hospitals		
Cant abort in the hospital	25	25
Out of hospital abortion does not lead to death	35	35
Even in the hospital abotees still die	30	30
A person will die when his time comes	7	7
No reason for inducing abortion outside hospitals	3	3
Total	100	100
Prevention of teenage pregnancies		
Parents to counsel young girls	13	13
Use professional counsellors	15	15
Improve family life education	35	35
Improve girl child parenting	4	4
Pay more attention to growing girl children	33	33
Total	100	100
Safeguards against teenage abortion		
Medical & counselling services	43	43
Recreational facilities	12	12
Preventive facilities	9	9
Free education	3	3
Parental socialisation	33	33
Total	100	100

Source: Author's Field Survey, 2011

An Efficient Heuristic Algorithm For Fast Clock Mesh Realization

P.Saranya, A.Sridevi

Abstract— The application of multiple clocking domains with dedicated clock buffer will be implemented. In this paper, an algorithm is proposed for determining the minimum number of clock domains to be used for multi domain clock skew scheduling. Non-tree based distributions provides a high tolerance towards process variations. The clock mesh constraints are overcome by two processes. First a simultaneous buffer placement and sizing is done which satisfies the signal slew constraints while minimizing the total buffer size by heuristic algorithm. The second one reduces the mesh by deleting certain edges, thereby trading off skew tolerance for low power dissipation by post processing techniques. Thus comparisons of wire length, power dissipation, nominal skew and variation skews using H-SPICE software for various sized benchmark circuits are performed.

Index Terms— Clock skew, Clock Distribution Network, Heuristic, Low Power Variation.

◆

I. INTRODUCTION

Clock skew refers to the relative difference of the clock latencies of registers. Clock skew is one of the design parameter that is very sensitive to process variations. Clock skew computes a set of individual delays for the clock signals of the registers and latches of synchronous circuits to minimize the clock period. In practice, a clock schedule with a large set of arbitrary delays is becoming unreliable. This is because the implementation of dedicated delays using additional buffers and interconnections is highly susceptible to within-die variations of process parameters.

The function of clock distribution network is to deliver the clock signal from the clock source to the clock sinks. In the formulation of the problem, process variation is not a direct consideration. The fact is, this technique helps to make it practical to inject a limited number of skew values into the circuit. A fast and efficient combinatorial algorithm is presented to design or optimize the distribution network and is also proposed that it can be used at design time to decide the number of distribution network to be used and their associated skew values such that the number of required buffers, as well as the resulting clock period, is minimized. Our proposal includes the following contributions:

- i. To find the mesh buffer locations and their sizes using greedy algorithm on a discrete set of libraries of buffer sizes.
- ii. The process of removing the noncritical wire segments in a clock mesh to minimize the power dissipated was to be formulated by network survivability theory.
- iii. We have to achieve the most desirable skew-power tradeoff, which becomes flexible enough to allow a high range of tradeoff.

Higher skew would bring down the maximum permissible delay. A mesh architecture is suited well for high-performance systems since it mitigates clock skew at the expense of high resource consumption. With power occupying an increasingly important role in chip design, it is necessary to find the most desirable skew-power tradeoff.

The remainder of this paper is organized as follows. Section II contains the list of Literature Survey done. Section III comprises of the methodology used in the work. Section IV gives the results and discussions. Section V gives the work done and conclusion.

II. RELATED WORK

In combinatorial optimization, the most important challenges are presented by problems belonging to the class NP-hard. The solution returned will be within a relative distance from the optimum. The design of approximation algorithm relies in a number of techniques, which usually involve clever analysis of the relaxation of integer problems to linear or semidefinite formulations. One of the motivations for developing approximation algorithms is to find good solutions for problems that are provably difficult. Another motivation is to better understand the intrinsic

P.Saranya, P.G.Scholar
Department of Electronics & Communication Engineering,
SNS College of Technology, Coimbatore, India
Email: digisaran@gmail.com
A.Sridevi
Department of Electronics & Communication Engineering,
Assistant Professor,
SNS College of Technology, Coimbatore, India
Email: srideviarumugam07@gmail.com

difficulty of the problems.[11] Link based non-tree clock distribution is a cost-effective technique for reducing clock skew variations. It is limited to unbuffered clock networks and neglected spatial correlations in the experimental validation. We overcome the shortcomings and make the link based non-tree approach feasible for realistic designs. The short circuit risk and multi-driver delay issues in buffered non-tree clock networks are examined. Skew tuning is used to synthesize a clock tree with low nominal skew under a higher order delay model. The effect of link insertion depends on a well-designed buffered clock tree which enhances the effectiveness of link insertion. Link insertion in a buffered network may result in multiple drivers for a subnet.[7]. Process variation, RLC delay matching and scalability with die size and process generation are the classic challenges in high speed clock distribution design. An architecture for achieving sub-10ps global clock uncertainty that addresses each one of these issues without additional clock jitter or layout area is presented. Included is a method to limit global clock skew to a single inverter stage delay independent of die size and management of practical constraints due to schedule, changing floor plan, die size, clock loading and process independence. Wirelength of clock routing trees should be minimized in order to reduce system power requirements and deformation of the clock pulse at the synchronizing elements of the system. The deferred-merge embedding (DME) algorithm, embeds any given connection topology to create a clock tree with zero skew while minimizing total wirelength. The algorithm always yields exact zero skew trees with respect to the appropriate delay model. The circuit speed is increasingly limited by two factors: i) delay on the longest path through combinational logic, and ii) clock skew, which is the maximum difference in arrival times of the clocking signal at the synchronizing elements of the design by the inequality governing the clock period of a clock signal.[3] A mesh construction procedure which guarantees Zero skew under the Elmore delay model, using a simple and efficient linear programming formulation. Buffers are inserted to reduce the transition time (or rise time). As a post-processing step, wire width optimization under an accurate higher-order delay metric is performed to further minimize the transition time and propagation delay skew[12]. Thus the hybrid mesh tree construction scheme can provide smaller propagation delay and transition time than a comparable clock tree.

III. PRELIMINARIES

Certain conventions and notations that are to be used in this paper are introduced below as

- i. Clock mesh has set of nodes with dimension $A_n \times B_n$. Buffer sizes ranges from 1 to K in a non-decreasing order. The i th buffer can drive a load capacitance of atmost C_i .
- ii. The buffer mapping functions maps each node location i and j to set of nodes and buffer sizes respectively. The problem implies that the i th value of buffer mapping function becomes null, thus it has no buffer in the i th location of the clock mesh.
- iii. The set of clock sinks ranges from S_1 to S_c without any loss of generality, thereby every buffer is connected to the set of nodes in the clock mesh. But these clock sinks are connected to the closest point in the mesh which are not available in the covering regions.
- iv. The minimum distance between nodes i and j in the mesh is given by A_{ij} . The maximum permissible delay, P_{ij}^{delay} between two registers i & j is given as, $P_{\text{delay}} = \min_{(i,j)} P_{ij}^{\text{delay}}$.

IV. HEURISTIC ALGORITHMS

A. Algorithm Description

A greedy algorithm is any algorithm that follows the problem solving heuristic of making the locally optimal choice at each stage with the hope of finding the global optimum. For example, applying the greedy strategy to the traveling salesman problem yields the following algorithm: "At each stage visit the unvisited city nearest to the current city". In general, greedy algorithms have five pillars:

1. A candidate set, from which a solution is created.
2. A selection function, which chooses the best candidate to be added to the solution.
3. A feasibility function, that is used to determine if a candidate can be used to contribute to a solution.
4. An objective function, which assigns a value to a solution, or a partial solution.

A solution function, which will indicate when we have discovered a complete Greedy algorithms appear in network routing. Using greedy routing, a message is forwarded to the neighbouring node which is "closest" to the destination. The notion of a node's location (and hence "closeness") may be determined by its physical location, as

in geographic routing used by ad-hoc networks. Location may also be an entirely artificial construct as in small world routing and table. A greedy algorithm can be thought of as a backtracking algorithm where at each decision point "the best" option is already known and thus can be picked without having to recurse over any of the alternative options. Greedy algorithms tend to be very efficient and can be implemented in a relatively straightforward fashion. Many a times in $O(n)$ complexity as there would be a single choice at every point solution.

Clock mesh implementation requires an array of mesh drivers, shown in fifth level of second block in fig1 is to drive the massive RC network of the clock mesh. The benefit of the mesh net is that it smoothes out the arrival time differences from the multiple mesh drivers that drive it. The top trace is the ideal clock, the top pair of traces shows the skew just before the mesh, and the bottom pair of traces shows the skew just after the mesh.

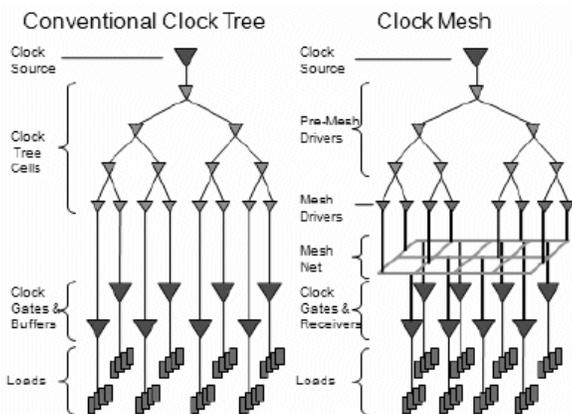


Fig1. Clock Structures - Conventional clock tree and clock mesh

B. Node Co-ordinate Determination

<p>Input :N,Number of buffers to be tested.</p> <p>Output:S,Set of Integrated nodes.</p> <ol style="list-style-type: none"> 1. Initialize S = null 2. Determine the mesh dimension $A_n \times B_n$ 3. Define the buffer sizes,K. 4. Calculate Effective capacitance $C_E = K/((CR - S)$ 5. Choose least $C_E \rightarrow LC_E$ 6. $S \leftarrow S \cup LC_E$
--

Fig2. Node co-ordinate integration for mesh by greedy

For the buffer mapping function to be estimated :
 1) the requirement of buffer is noted. 2) the size of buffer needed with the each node in the mesh allocated to at least one buffer and each buffer driving less than the maximum load it can drive. The Covering Region(CR) of the node for a particular buffer is defined as the set of nodes around the node in the 2-D mesh such that the total capacitance of the nodes included in the covering region is less than the maximum capacitance that the buffer can drive. If any more edges in the mesh are added, then the capacitance of the region will be greater than the maximum load that the buffer can drive. The greedy algorithm pick the set that covers the most nodes and then throw away the nodes that are covered. The process is repeated until all nodes are covered.

The algorithm may return two buffers for the same location, which is not a feasible solution. Thus, such a situation can be easily avoided by using the observation.

Observation 1: A bigger buffer size can drive a bigger load. For any node i with solution Φ , if there are more than one buffer driving a node, one can pick the biggest buffer without losing feasibility.

C. Steiner Reduction

After the integration of the position of the mesh buffers and their sizes ,the next task is to reduce the size of the mesh. This is done by removing edges such that a certain level of redundancy is still maintained. Remove edges from the mesh such that 1)each sink S_i has at least k node locations such that for each such node locations j, $A_{ij} \leq L_{max}$, $BM(j) \neq \Phi$ and there exists at least l edge disjoint paths between j and i and 2) the number of edges removed is maximized.

<p>Input : The buffer cost and node to be tested.</p> <p>Output : B_R , Reduced buffer cost.</p> <ol style="list-style-type: none"> 1. Define the cost of node, e. 2. Initialize $B_R \leftarrow \Phi$ 3. Find K nearest nodes for the tested nodes. 4. Determine the minimum cost , I_{min} 5. For each $e \square I_{min}$ 6. $B_R \leftarrow B_R \cup e$
--

Fig3. Steiners Reduction

The user defined parameters control impact the solution as k and l high would mean more redundancy and hence more tolerance to variations but less number of edges removed or more power dissipation. Thus after mesh reduction the buffers could be driving a load that is significantly lesser than that of a complete mesh. Hence, as a postprocessing step we compute the new load and downsize the buffers accordingly.

V EXPERIMENTAL RESULTS

In order to mention the effectiveness of the aforementioned techniques, we conducted the following algorithms 1) Compare our node co-ordinate integration greedy algorithm for various sizings. We do the comparisons for the minimum, medium and maximum sizing of buffers, 2) Comparison of our Steiner reduction algorithm with that of the complete meshes, 3) Finally the run time results are also achieved and compared against various sizing of buffers in a complete mesh.

The greedy and Steiner reduction algorithms were implemented in C++ and the simulations were generated on a Linux work station .The test cases of the benchmark circuits were simulated in HSPICE using 180nm process model from Berkeley Predictive Technology Model and the results are compared for various sizing of circuits.

approximately while the medium and maximum sizing achieves with a larger power as obtained in waveform simulation.

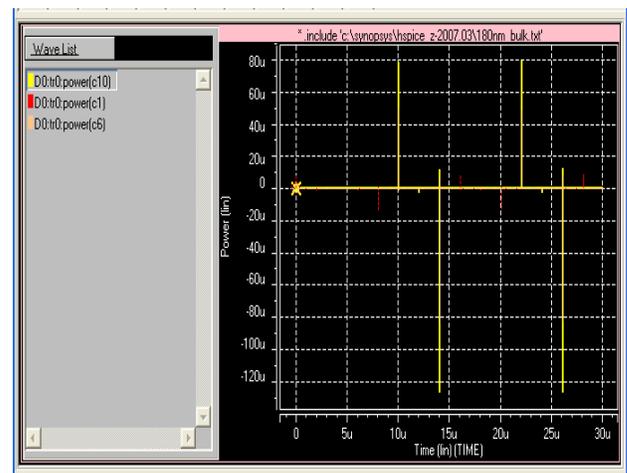


Fig 5. Output waveform with medium size buffer

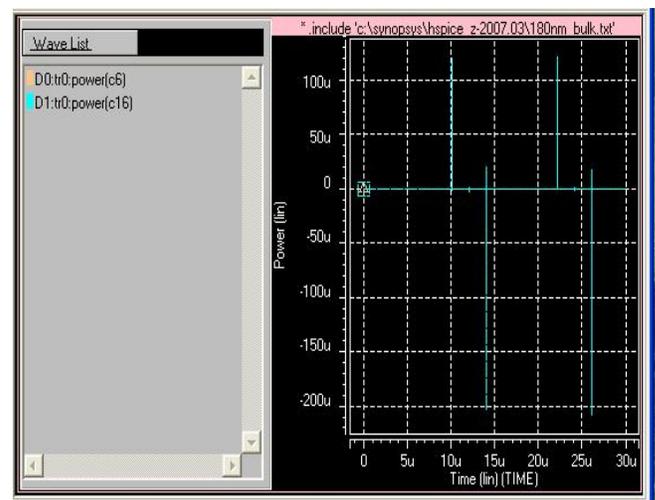


Fig 6. Output waveform with maximum size buffer

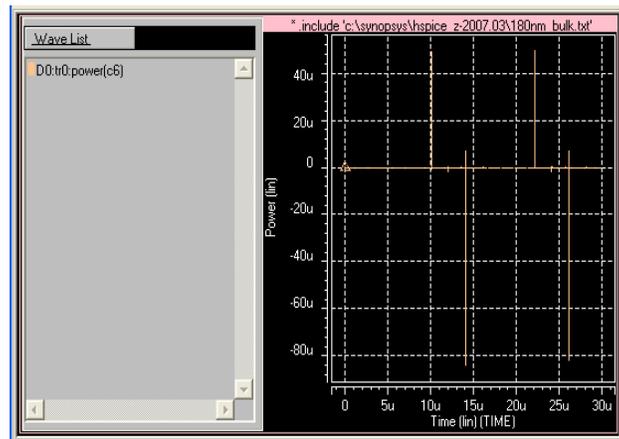


Fig 4. Output waveform with minimum size buffer

The run time of our algorithm is within a few seconds as verified from the Table1. It becomes one of the fine tuning technique and not an optimizing procedure. Such a reduction in area though small could be significant in high performance designs. Hence the minimum buffer sizing achieves the low power dissipation of about 40u

Thus the Fig4 ,Fig5 and Fig 6 gives the power versus clock time trade-off curve for test case S9234 .The Y-axis measures the power dissipation and X-axis measures the time value for the various sizing of the circuits done using heuristic algorithms by a HSPICE model. The Table1 provides the comparison of parameters (memory,cputime,etc..) for various sizings. Thus the following inferences can be made as Using the minimum buffer size results in an area reduction of about 10% comparitively to other sizings. Medium buffer sizes satisfy slew constraints in all but except of one case whereas large buffer size satisfies the slew constraints in all the cases. But it can be achieved only with an area penalty of 18%(35%) for medium (maximum)buffer sizes.

TABLE 1

RESULTS FOR BUFFER SIZING FOR TEST CASE S9234

Test case – S9234	Sizing		
	Minimum	Medium	Maximum
Memory used	1097 kbytes	1865 kbytes	3547 kbytes
CPU Time	3.8 sec	6.48 sec	14.01 sec
Transient time	2.65	5.67	12.96
Mosfets defined	150	290	560
Total Iteration	9900	10894	12130
Output analysis	0.39	0.66	0.86
Operating point	0.05	0.07	0.11
Nodes	141	273	531
Elements	308	592	1138

VI CONCLUSION

The research work has been focused on sizing the interconnect elements within the clock mesh. We believe that sizing the interconnect wire segments of the clock mesh and the buffers driving the mesh simultaneously would yield more improvement in the power of a clock mesh area satisfying the constraints. The work presented here can be easily extended for sizing buffers and mesh elements simultaneously. Since, our design techniques are faster, it offers the flexibility to optimize clock mesh with different design objectives. The slew constraints are also satisfied while minimizing the total buffer size in a simultaneous buffer placement and sizing module. The number of mesh in the distribution network is also reduced by deleting certain edges, thereby trading off skew tolerance for low power dissipation. The various benchmark circuits are considered and their performance measures are determined by above heuristic algorithms.

REFERENCES

- [1] J. Burkis, "Clock tree synthesis for high performance ASIC's," in *Proc. IEEE Int. Conf ASIC*, pp. 9.8.1-9.8.4, 1991.
- [2] P. K. Chan and K. Karplus, "Computing signal delay in general RC networks by tree/link partitioning," *IEEE Trans. Computer-Aided Design*, pp. 898-902, Aug. 1990.
- [3] S. Dhar, M. A. Franklin, and D. F. Warm, "Reduction of clock delays in VLSI structures," in *Proc. IEEE Int. Conf Computer Design*, pp. 778-783, 1984.
- [4] A. Rajaram, D. Pan, and J. Hu. Improved algorithms for link based non-tree clock network for skew variability reduction. In Proceedings of the ACM International Symposium on Physical Design, pages 55(62, 2005).
- [5] M. Edahiro. A clustering-based optimization algorithm in zero-skew routings. In Proceedings of the ACM/IEEE Design Automation Conference, pages 612(616, 1993).
- [6] S. Pullela, N. Menezes, J. Omar, and L. T. Pillage. Skew and delay optimization for reliable buffered clock trees. In Proceedings of the IEEE/ACM International Conference on Computer-Aided Design, pages 556(562, 1993).
- [7] J. G. Xi and W. W.-M. Dai. Buffer insertion and sizing under process variations for low power clock distribution. In Proceedings of the ACM/IEEE Design Automation Conference, pages 491(496, 1995).
- [8] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy, *Proof verification and the hardness of approximation problems*, J. ACM, 45 (1998), pp. 501–555. Prelim. version in FOCS'92.
- [9] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, And M. Protasi, *Complexity and approximation*, Springer-Verlag, Berlin, 1999. Combinatorial optimization problems and their approximability properties, With 1 CD-ROM (Windows and UNIX).
- [10] U. FEIGE, *A threshold of $\ln n$ for approximating set cover (preliminary version)*, in Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing (Philadelphia, PA, 1996), New York, 1996, ACM, pp. 314–318.
- [11] *Approximation thresholds for combinatorial optimization problems*, in Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002), Beijing, 2002, Higher Ed. Press, pp. 649–658.
- [12] S. Pullela, et al., "Skew and Delay Optimization for Reliable Buffered Clock Trees," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 556-562, 1993.
- [13] J.P. Fishburn, "Clock skew optimization," *IEEE Transaction on Computers*, vol. 39, pp. 945-951, Jul. 1990.

Factors Affecting Acceptance of HIV Counseling & Testing Among Antenatal Care Attendants: With Emphasis on Role of Male Partners

Addisu Belachew (BSc., MPH/RH), Abebe G/Mariam (BSc., MPH/RH, Professor)

Abstract— VCT is an entry point for Prevention of mother to child transmission of HIV infection. To increase uptake of Prevention of mother to child transmission interventions and to benefit more generally from HIV testing, the greater role of men is important. Facility based cross-sectional by design, conducted from March 1 to April 1, 2011. The sample size was 422 and study subjects were selected by systematic random sampling. Acceptance of HIV testing among the interviewed pregnant women was (72.0%). Husband reaction, fear of rejection by the community and fear of positive test result were reasons reported that impede acceptance of HIV testing.

Index Terms— antenatal care visit, breast feeding as a means of MTCT, use of drug & avoid breast feeding as PMTCT, previous testing, disclose test results to their partners, partner approval & partners would accept invitations are predictors.

◆

1 INTRODUCTION

An estimated 430 000 new HIV infections occurred among children under the age of 15 in 2009 globally. Most of these new infections are believed to stem from transmission in utero, during delivery, Post partum as a result of breast feeding. Vertical mother-to-child transmission accounts for an estimated 91% of pediatric HIV cases in sub-Saharan Africa [1]. It is estimated that without any intervention about 35% children born to HIV-infected mothers will be infected with the virus (2). This percentage has reportedly been reduced to levels as low as 2% in developed countries with the implementation of core PMTCT interventions and full uptake of HCT among ANC attending pregnant women [3], [4].

However, low uptake of HIV testing by antenatal women is challenging achievement of global commitment to reduce the proportion of infants infected with HIV and effectiveness of PMTCT programs in sub-Saharan Africa [5].

In Ethiopia, nowadays PMTCT programs being integrated within ANC service in health institutions. During ANC service, pregnant women are offered group counseling. It was found that however that after counseling there are many women who don't get tested the same day (about 40%). women are affected by different factors that can restrain them from HCT. PMTCT intervention is strongly focused on pregnant women who are usually tested alone, often without involving their partners. Lack of involvement of partners has been suggested as a reason for the low uptake of HCT and PMTCT interventions in some countries reports and studies [6], [7]. On national guideline of PMTCT program, the low level of male involvement was mentioned as a potential gap that indicates low program effectiveness [8]. In addition on Ethiopia Epidemiological Synthesis of HIV/AIDS, low men involvement was considered as one of the main challenges to increase access and provide quality PMTCT services [9].

Studies in Ethiopia and other parts of the world indicate

that HIV testing acceptance is encouraging to improve the PMTCT uptake. Stigma and discrimination, husbands' negative reactions and fear of positive test result were reasons that impede higher acceptance of the test [5], [10], [11]. Majority of the pregnant women do not decide independently for acceptance of HIV testing. Decision-making authority is commonly referred to their male partners. Male partners' involvement in HIV testing of pregnant women was found low. Pregnant women need their male partners' positive attitude and support to accept HIV testing [10].

In the Ethiopian women traditionally are under influence of men and there is power imbalance between men and women. This could have great implication to acceptance of HIV testing and PMTCT program. There is limited data on the rate and role of male partner in acceptance of HCT uptake among pregnant women. Hence this study has assessed factors affecting acceptance of HIV testing among ANC attendees; described decision making of pregnant women on HIV testing and tried to indicate strategies to enhance the contribution of male partners in PMTCT services.

2 Methods and Materials

2.1 Study area and period

The study was conducted at health facilities which were providing both PMTCT & HIV counseling and testing in East Gojam Zone from March to April 2011, in five health centers: Tsedemariam, Yehidasew, Lega, Amber and Lumame health centers.

2.2 Study Design

Facility based cross-sectional study using mixed method

2.3 Study population

For Quantitative Study: Sampled pregnant women attending ANC who received HIV pre-counseling and testing in those public health centers.

For Qualitative Study: Purposive sample male partners, any pregnant women, health workers who were directly in-

involved in the provision of PMTCT/HCT service, woreda health office experts, zonal health department experts & NGO experts.

2.4 Sample size calculation

The sample size was determined by using single population proportion formula by assuming: P the proportion of pregnant women who undertake joint decision to be 50%, d margin of error (5%), $Z_{\alpha/2}$ was critical value at 95% confidence level of certainty= 1.96, $n = \frac{Z_{\alpha/2}^2 [p(1-p)]}{d^2}$ adding non-response rate of 10%, the required sample size was 422. A total of four FGD sessions were conducted (a total of 39 participants); two among pregnant women, two among male partner and 12 key informants.

2.5 Sampling procedure

The number of study units to be sampled from selected public health centers was determined using proportional allocation to size based on the number of client flow in the previous year of total average one month case load. Systematic random sampling was employed to select and approach each study subjects.

2.6 Data collection instrument

Structured interviewer administered questionnaire was used to collect the data which was adapted from different literatures, similar studies and modified according to the local context & objectives of this study. For both FGD & in-depth interview, semi structured discussion and interview guide was prepared. Client exit structured interviewer administered questionnaire interview for pregnant women at ANC clinics. Before the actual data collection, the quantitative questionnaire was pre-tested on 5% of the total sample size outside the study area. Ten trained diploma level clinical nurse graduates from medical college, two BSC nurse supervisors and two facilitators were recruited and participated throughout the data collection.

2.7 Data processing & analysis

For quantitative data, after data collection, each questionnaire was checked for completeness and code was given before data entry. Data was entered, cleaned, explored for outliers, missed values and missed variables and analyzed using SPSS version 16 statistical packages. Bivariate analysis was performed and then those variables that showed significant association with the outcome variable were included in a single model and multiple logistic regressions were conducted.

For the in-depth interview & FGD, after the interview data was transcribed word by word into the Amharic language and then translated into English language. Then similar ideas was grouped and summarized based on the key variables of the study.

2.8 Ethical Consideration

The study was obtained Ethical clearance from concerned authorities and offices. Similarly after clear discussion about the actual study written informed consent was obtained from each study subjects while the study subjects right to refuse was also be respected.

3 RESULTS

Socio-demographic characteristics

A total of 414 ANC attendants were participated in this study making the response rate of 98%. The mean age of respondents was 27 years with SD of 6.33. Most of the respondents, 380(91.8%), are currently married. Among the total respondents, 390(94.2%) are Orthodox Christians in religion. Four hundred thirteen (99.8%) respondents are Amhara in ethnicity. Greater than half of the respondents 223 (53.9%), are house wives, 83 (20.0%) are government employers and 53 (12.8%) are merchants by their occupation. The mean age of respondents' husband was 34 years with SD of 7.96. About (45.5%), (39.9%) of pregnant women and their partners had formal education respectively.

Knowledge of ANC Attendees on MTCT and PMTCT

Three hundred twenty one (77.5%) knew that a mother with HIV can pass the virus to her baby. Eighty four (20.1%) and forty two (10.1%) knew at least two and all the three mode of transmission (pregnancy, labor, and breast feeding) of MTCT respectively. About (24.4 %) didn't know what measures a HIV positive pregnant woman could take to prevent MTCT; the method of prevention of MTCT mentioned by the respondents: 182 (44.0%) use of drug, 202 (48.8%) avoid breast-feeding, 46 (11.1%) safe delivery service a means of PMTCT. Two hundred forty nine (60.1%) of the pregnant women knew the existence of PMTCT at the health facilities and 165 (39.9%) don't know the existence of the program.

HIV Testing Acceptance of pregnant women and HIV Counseling & Testing Services

Three hundred fifty three (85.3%) respondents had heard HIV counseling & testing. One hundred thirty (31.4%) of the respondents heard the presence of the service from multiple sources (mass media and health workers). HIV testing acceptance was found to be 298 (72.0%). Respondents were asked to mentioned some of the reasons for refusal of HIV counseling & testing at ANC setting: (84.3%) of pregnant women were answered deal with stress of being positive; (45.2%) mentioned uncertainty about husbands reaction; (63.5%) fear of rejection by the community; (32.1%) due to non respect of confidentiality.

Decision Making of the ANC Attendees towards Acceptance of HIV Testing

Regarding decision making towards acceptance of HIV testing, one hundred forty six (35.3%) of pregnant women responded that they can decide independently, but 64.7% need their partner's involvement. Three hundred forty seven (83.8%) pregnant women have responded on the important of partner consultation prior to HIV testing but about sixty six (16.2%) of respondents did not need partner consultation. Some of their reasons were; (31.9%) said their partners always negative for them, (30.4%) responded they did not have a habit of discus-

sion, (24.6%) mentioned their partners was not willing to consult them and only (13.1%) said it is due to culture. Health center staffs key-informants agreed that majority of pregnant women need consultation of their partner or nearby relatives. One of the center staffs key-informants said, "...As to me the problem is community culture & value, even if our women accept HIV testing, most of them need consultation of their partners or nearby relatives.....without their partners consult anything they do.....just to respect community norms and cultures..."

Role of male partner involvement in HIV testing of pregnant women

Only sixty five (15.7%) women were accompanied by their male partners to the health centres for ANC or HIV testing, but (84.3%) came alone or accompanied by relatives/friends. Some of the reasons why male partners not accompany with pregnant women at ANC clinic mentioned by the respondents were; men's would be overloaded with other works, pregnancy related services are considered as the task of pregnant women only, men's were not willing to go with us, fear of stigma & discrimination and fear of positive test result. As shown below fig. 1.

One of the key informants from zonal health department said "... antenatal care as a modern health care practice is a shared domain between husbands and their pregnant wives. Male partners followed cultural practices and norms regarding antenatal care because of traditional norms expected from them in the society and due to this in this area accompany of male partners with pregnant women not as such common..."

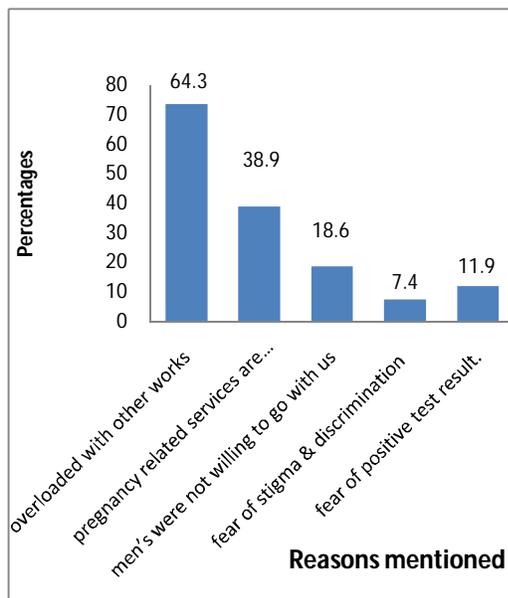


Figure 1: Reasons mentioned by the respondents why male partners not accompany with pregnant women at East Gojam public health centers ANC clinic, North-West Ethiopia, April, 2011

The women were asked whether their partner would

come for HIV testing in their next visit. Two hundred seventy (65.2%) of the participants said that their partner may accompany them in the next visit, but the rest said their partners would not be willing. Some of the reasons were mentioned for refusing to join during the next visit; their partners are highly busy with different affairs, don't have a habit of open discussion, males don't have a practice of accompany with them at ANC clinics, fear of stigma & discrimination and fear of positive test result. The respondents' reasons were also frequently reported by male FGD participants not to accept the invitation. One of the participants said: "many male partners don't think that pregnancy related services are the task of men's due to cultural norms. The other reasons, fear of positive test result and community feeling for HIV/AIDS are major reasons, for example I am one of them to think like this..... As to me testing of HIV is like a death penalty.... If I convince myself to go, it would be a bad decision for me due to fear of positive test result, but now I don't believe (have the courage)..."

Accordingly three hundred nine (74.6%) said that they would get approval but the rest one hundred five (25.4%) did not get approval. Some of the reasons mentioned were: (56.2%) of participants answered their partners were not willing; (27.6%) of pregnant women said did not ask their partners; (16.2%) of participants responded did not have a habit of discussion. Only two hundred nine (50.5%) had a habit of discussion on HCT and other health issues with their partner. Almost half of the study participants (49.5%) have no habit of discussion on HCT and other health issues with their partners.

Only (44.2%) pregnant women answered, their partners would help if their test result was found to be positive, but (32.4%) women responded that their partner might not sure them support if their test result was found to be positive and (23.4%) women responded that their partner would render them support if their test result was found to be positive. Most women in the focus groups regarded their husbands as difficult to deal with; a number of reasons for involving them were mentioned. These were mainly related to women's low decision-making status relative to men. Women were concerned about disclosing their HIV status to a partner who had not been tested.

Factors affecting acceptance of HIV testing among pregnant women

Association with acceptance of HCT rested with number of ANC visit, breast feeding as MTCT, use of drug & avoid breast feeding as PMTCT, prior HIV testing HCT, disclose HIV test result to your partner, partners' approval and those partners would accept invitation. The odds of Pregnant women who had got partners approval were 4 times to be tested than the odds of those who had no got partners approval (AOR= 3.85, 95% CI= (1.63, 9.07) and the odds of women who had prior HIV testing were 6 times to accept HCT than the odds of those who had no prior HIV testing (AOR= 5.8, 95% CI= (2.09, 16.25)

Table 1: Adjusted determinant factors of accepting HCT among ANC attendants at East Gojam public health centers ANC clinic,

North-West Ethiopia, April, 2011

Variables	Acceptance of HIV counseling & testing			
	Yes Num (%)	No Num (%)	COR(95% C.I)	AOR(95% C.I)
ANC visit				
First visit	158(53.0)	105(90.5)	0.1(0.03-0.48)*	0.23(0.05-0.75)**
Second visit	113(37.9)	9(7.8)	0.9(0.2-4.5)	1.1(0.4-7.6)
Third & more	27(9.1)	2(1.7)	1	1
During post-partum period				
Yes	248(83.2)	46(39.7)	7.5(4.7-12.2)*	3.1(1.2,8.0)**
No	50(16.8)	70(60.3)	1	1
Use ART				
Yes	169(56.7)	13(11.2)	10.4(5.6-19.3)*	.44(1.7-11.5)**
No	129(43.3)	103(88.8)	1	1
Avoid breast feeding				
Yes	185(62.1)	17(14.7)	9.5(5.4-16.8)*	3.1(1.2-8.0)**
No	113(37.9)	99(85.3)	1	1
Ever had HCT				
Yes	185(62.1)	13(11.2)	12.9(7.0-24.1)*	5.8(2.1,16.2)**
No	113(37.9)	103(88.8)	1	1
Disclose test result				
Easy	251(84.2)	25(21.6)	19.4(11.3-33.4)*	16.7(6.9,40.3)**
Difficult	47(15.8)	91(78.4)	1	1
Partners approval				
Yes	251(84.2)	58(50.0)	5.3(3.3-8.7)*	3.8(1.6,9.1)**
No	47(15.8)	58(50.0)	1	1
Partners invitation				
Yes	213(71.5)	57(49.1)	2.5(1.7-4.0)*	2.5(1.1,5.8)**
No	85(28.5)	59(50.9)	1	1

4 DISCUSSION

HIV testing in pregnancy is the gateway to accessing care for the mother and the child. The intervention in prevention of mother-to-child transmission (PMTCT) of HIV can only be applied to a woman whose status is known. Therefore determining the HIV status of pregnant women is a key factor to the success of any prevention program.

This study revealed that HIV testing acceptance among ANC attendees was found to be 72.0%. However it is higher than the 2009 national coverage (67.7%) [12]. The difference between the study finding and the performance report of national level could result in underreporting of the health facilities. But it is lower than studies done from Ethiopia, Ghana and the Ethiopian government target which was set at 80% [6], [13], [14]. The possible reason why this finding is lower than other studies may be the accessibility & availability of IEC/BCC materials about MTCT/PMTCT & the socio-demographic characteristics difference of the current study respondents versus the other studies respondents. This implies that it is encouraging to increase PMTCT uptake in the area and more effort should be made to increase the acceptance rate. The most common reasons to refuse testing were need to discuss with partner, fear of HIV positive status, fear of loss of marital security, domestic violence and confidentiality.

The finding of this study also showed significant association between the number of antenatal visit and acceptance of prenatal HIV testing. This finding is consistent with studies conducted in Ethiopia, Adigrat and AA [13], [15]. This association between number of antenatal visit and acceptance of prenatal HIV testing may be explained by frequent exposure of mothers to information regarding HCT, MTCT and PMTCT during their follow up, which may influence the mother to take the test.

Respondents knowledge of PMTCT, those mentioned use of drug and avoid breast feeding were predictors of acceptance of HCT. Studies done in Ethiopia, Wolaita and Mekelle, knowledge about treatment that reduces mother-to-child transmission of HIV was found independently associated with testing [10], [14]. This finding also strengthens the fact that awareness on PMTCT is critical in increasing acceptance of HIV testing.

This study also revealed that the odds of pregnant women who had HIV testing in the past were six times to be tested than the odds of those women who had no prior HIV testing experience (AOR= 5.8, 95% CI= (2.09, 16.25). When adjusted for other variables prior HIV testing experience was independently and significantly associated with HCT acceptance. This finding was also reported by other studies. Studies conducted in Addis Ababa, Arba-minch Ethiopia, Barbados and Zambia indicated that having HIV test in the past was an independent factor positively influencing readiness for testing [16], [17], [18]. [19]. The possible explanation for association between previous testing and current one is that women who had HIV testing in the past are more likely to have change in their sexual behavior after knowing their serostatus.

Pregnant women who replied that it is easy to disclose test results were associated with acceptance of HCT among pregnant women. This finding similar with a study done in wolaita [10]. The reasons respondents mentioned that why it is difficult to disclose test result were divorce, physical violence, rejection and blame.

Sixty five (15.7%) of pregnant women were accompanied by male partners for ANC/ HIV testing. The result of this study is higher than studies done in wolaita which is 5.1% and AA it was less than 10% [6], [10]. But it is lower than the 25% level set by HAREG [6]. It is consistent with a study done in Kenya (16%) [20]. The finding of this study also showed pregnant women were associated with acceptance of HCT but not statistically significant. The study done in Tanzania also indicated that pregnant women were less likely to accept HIV testing if their partners did not accompany them [21]. A study done in wolaita also reported that, those pregnant women accompanied by their partners were 1.3 times higher chances to accept HCT than if their partners did not accompany them [10]. This reveals that it is highly to attract male partners to accompany spouses to ANC clinic sessions at least once during pregnancy. Key-informants

explained that a lack of community awareness about the importance of partner HCT and cultural beliefs that men should not participate in antenatal activities may explain low partner involvement. National and local VCT campaigns promoting couple counseling can be used to address barriers to testing male partners and increase awareness in the general public.

Women discussants said that, if a woman is HIV tested, but her husband does not agree to be tested, he has to be counseled about prevention. This will call men's attention to the problem and may make them more responsible. There is not much use in counselling and HIV testing women only. It is really necessary that men also be involved. One of the barriers was that men were not regarded as responsible for issues relating to pregnancy and childbirth. One of key-informants from the health facilities said that "... *If men were approached, there would be a chance to increase their involvement...*"

In this study 35.3% of pregnant women can independently decide to accept HIV testing. The figure is higher than studies done in India which is only 21% and Ethiopia, wolaita which is 27% [10], [22]. This difference may be due to time series of the studies conducted and the respondents of the current study improves of their decision making power through time & the other reason may be difference of the studies participants, the power and position they are given at household or community level. This can be explained by the fact that women hold lower social and economical positions than men and socially constructed aspects of male and female relationships within societies influenced the different health outcomes of women and men. Furthermore the qualitative part of this study showed that women's ability to decide for any health seeking behavior is determined by the power and position they are given in household. PMTCT interventions may be complicated by women's lack of decision-making authority, which is commonly deferred to their male partners; hence male partners appear to play an important role on women's decisions to accept HIV testing.

Partners test approval was significantly associated with acceptance of HIV testing among pregnant women. This finding is similar with a study done in Gambella region reported that women who thought their husbands would approve were almost six times to test than those who thought their husbands would not approve (AOR= 5.6, 95% CI= 2.8, 11.2) [23]. The reasons of respondents those who didn't get partners approval were partners not willing, they did not have open discussion and did not ask them at all before testing of HIV.

5 Conclusion

HIV testing acceptance is encouraging to improve the PMTCT uptake. But, stigma and discrimination, hus-

bands' negative reactions and fear of positive test result were some of the reasons that impede higher acceptance.

Two third of pregnant women do not decide independently for acceptance of HIV testing. Decision-making authority is commonly referred to their male partners. HIV testing acceptance is encouraging but men's involvement was found to be low. As a result of fear of rejection by community & deal with stress of being positive test result, large number of male partners didn't accompany with their partners at ANC clinics. Pregnant women need their male partners' positive attitude & support to accept HIV testing.

Recommendation

_Efforts should be given for intensive and continued information dissemination, to both pregnant mother and their partners, about prenatal HIV transmission, the role of HIV counseling and testing (HCT) on the prevention of mother-to-child transmission of the virus, and about the existence of intervention that reduce the possibility of prenatal transmission of the virus.

_Efforts must be made to achieve full attendance of ANC by all pregnant mothers. For this to be realized effective health promotion programs need to be emphasized.

_Health centers should use couple counseling as a strategy to improve male involvement;

_Couple counseling facilitated through couple-friendly ANC services could be taken as a strategy to minimize the difficulty that pregnant women face to disclose their HIV test result to their partner;

_Strengthen VCT within the health centers as well as outreach VCT sites;

_Counseled (tell) pregnant women's to invite their partners in the next, if they don't accompany with their partners;

_Male friendly counseling should be strengthened at PMTCT institutions.

_Regional health bureau should make the IEC/BCC activities of PMTCT tailored to improve male involvement;

_Organizations working on PMTCT program should give emphasis on involvement male partners;

ACKNOWLEDGEMENT

I would like to thank my advisor Prof. Abebe G/M for his valuable advice and comments they have provided during the development of this project research thesis.

REFERECES

[1] Joint United Nations Program on HIV/AIDS (UNAIDS) and World Health Organization (WHO). Global AIDS epidemic update Report, November 2009.

[2] UNAIDS. UNAIDS Best Practice of Pediatric HIV infection and AIDS, September 2002.

[3] Newell M. Prevention of mother-to-child transmission of HIV; challenges for the current decade. Bulletin of the world Health Organization.2001; 79(12):1138-1144.

[4] Phoolcharoen W, Detels R. Acquired immunodeficiency syndrome. In : Oxford of Public health, edited by

R Detels, J McEwen, R Beaglehole and H Tanka. 4th edition. 2002; New York; Oxford University Press.

[5] Cartoux M, Meda N, Van de Perre P et al. Acceptability of voluntary HIV testing by pregnant women in developing countries: an international survey. Ghent international journal, March, 2005. F. Baiden et al. Antenatal-linked voluntary counselling working group on mother-to-child transmission of HIV. AIDS, 1998; 12: 2489-2493.

[6] Takele A. Assessment of male partner influence on pregnant women towards voluntary HIV testing and support on PMTCT in Hospitals of Addis Ababa, MPH thesis, AAU, 2007.

[7] Theresa M., Exner N. et al. HIV Counseling and Testing: Women's Experiences and the perceived Role of Testing as a prevention strategy. On Sexual and Reproductive Health article, March/April, 2002; Volume 34 (2): Nigeria.

[8] Ministry of Health. National Guideline on the Prevention of mother-to-child transmission of HIV in Ethiopia, Addis Ababa, November 2007.

[9] HAPCO and GAMET. HIV/AIDS in Ethiopia Epidemiological Synthesis, April 2008.

[10] Tilaye T, Factors Affecting Acceptance of HIV Testing Among Antenatal Care Attendees: With Emphasis On Role of Male Partners. EPHA Sponsored Master's Thesis Extracts on HIV/AIDS, January, 2010; Extract No. 9, Wolaita Zone, SNNPR, Ethiopia.

[11] Paulos M, Amaha H, Sileshi L, Ayele B. Determinants Of VCT Uptake Among Pregnant Women Attending Two ANC Clinics. Ethiop Med J, 2007; 45 (4), Addis Ababa, Ethiopia.

[12] Federal Democratic Republic of Ethiopia Ministry of Health. Health and Health related Indicators. Policy Plan Directorate Monitoring and Evaluation Case Team, AA, Ethiopia, 2008/2009.

[13] Samuel T. Assessment of factors contributing to utilization of provider initiated HIV testing and Counseling among Women attending ANC in Addis Ababa. EPHA Sponsored Master's Thesis Extracts on HIV/AIDS, 2008; Extract No. 12, AA, Ethiopia.

[14] Atsbha G/K. Factors affecting acceptance of voluntary HIV Counseling & Testing among pregnant women Attending Antenatal Care in Mekelle Public Health Facilities. EPHA Sponsored Master's Thesis Extracts on HIV/AIDS, Feb.-March 2009; Extract No. 11, Mekelle, Ethiopia.

[15] Tesfaye G. Factors determining acceptance of VCT for PMTCT among urban and rural pregnant mothers in Adigrat and Ganta-Afeshum woredas. EPHA Sponsored Master's Thesis Extracts on HIV/AIDS, Nov-Feb 2008/09; Extract No. 12, Tigray-Ethiopia.

[16] Temesgen K. Assessment of the Determining Factors for Acceptance of HIV Testing among Pregnant Women at Antenatal Care Setting. EPHA Sponsored Master's Thesis Extracts on HIV/AIDS, January, 2010; Extract No. 9, Arbaminch Town, SNNPR, Ethiopia.

[17] Knut F. and Seter S. A randomized trial on acceptability of voluntary counseling and testing Tropical medicine and international health, May, 2004; 9(5):566-572.

[18] Alok K., Elizabeth R., et al. Antenatal voluntary counseling and testing for HIV in Barbados. Success and barriers to implementation. Pan Am J public health, 2004, 15(4):242-247.

[19] Getachew W. Factors Determining Acceptance of Voluntary HIV Testing Among Pregnant Women Attending Antenatal Clinic At Armed Force Hospitals In Addis Ababa. AA University Master's Thesis, unpublished, 2005, AA, Ethiopia.

[20] D A Katz, J N Kiarie, et al. HIV testing men in the antenatal setting: understanding male non-disclosure. International Journal of STD & AIDS, January, 2009; 20: 765-767, Kenya.

[21] S. E. Msuyaabc. M. Mbizvoad, et al. Low male partner participation in antenatal HIV counseling and testing: implications for preventive programs. AIDS Care, July, 2008; Vol. 20 (6): 700_709, Moshi urban district, northern Tanzania.

[22] H Brown, et al. Attitude towards prenatal HIV testing and treatment among Pregnant women in southern India. International Journal of STD & AIDS, 2001; 12:390-394.

[23] Wondimagegn F. Assessment of Determinants for Refusal of HIV Counseling and Testing Service Utilization among ANC Attendees. EPHA Sponsored Master's Thesis Extracts on HIV/AIDS, January, 2010; Extract No. 9, Gambella, Ethiopia.

Table 1: Socio-demographic characteristics of ANC attending pregnant women at East Gojam public health centers, North-West Ethiopia, April, 2011

Socio-demographic characteristics	Number	Percent (%)
-----------------------------------	--------	-------------

Age (year)		
15-24	144	34.7
25-34	194	46.9
35-45	76	18.4
Residence		
Urban	188	45.4
Semi urban	40	9.7
Rural	186	44.9
Educational level		
No formal education	225	54.5
Formal education	189	45.5
Number of pregnancy		
Primipara	156	37.7
Multi Gravida	203	49.0
Grand Multipara	55	13.3
No. of ANC visit		
First visit	263	63.5
Second visit	122	29.5
Third & more visit	29	7.0
Marital Status		
married	380	91.8
unmarried but in stable union	34	8.2
Respondents' occupation		
Employed in any organization	83	20.0
Unemployed	331	80.0
Husband's Occupation		
Employed in any organization	83	20.0
Unemployed	331	80.0
Husband's educational level		
No formal education	249	60.1
Formal education	165	39.9
Husband's Age in complete years		
21-30	181	43.7
31-40	164	39.6
41-50	55	13.3
51-60	14	3.4
Family monthly income		
<500 birr	298	72.0
>=500 birr	116	28.0
Length of years in relationships		
< 5	163	39.3
5-10	133	32.1
10-15	74	17.9
> 15	44	10.6
Total	414	100

Table 2: Association of socio-demographic characteristics and acceptance of HIV counseling and testing among ANC attendants at East Gojam public health centers ANC clinic, North-West Ethiopia, March – April, 2011

Variables	Acceptance of HIV counseling & testing		
	Yes Num (%)	No Num (%)	Crude OR (95% C.I)
Residence			
Urban	172(57.7)	16(13.8)	9.6(5.4,17.4)*
Semi urban	28(9.4)	12(10.3)	2.1(1.0-4.4)*
Rural	98(32.9)	88(75.9)	1
Educational level			
Formal education	170(57.0)	19(16.4)	6.8(3.9,11.7)*
Informal education	128(43.0)	97(83.6)	1
Respondents' occupation			
Employed in any organization	77(25.8)	6(5.2)	6.3(2.7-15.1)*
Unemployed	221(74.2)	110(94.8)	1
Family monthly income			
< 500 birr	194(65.1)	104(89.7)	0.2(0.1-0.4)*
>= 500 birr	104(34.9)	12(10.4)	1
No. of ANC visit			
First visit	158(53.0)	105(90.5)	0.1(0.03-0.48)*
Second visit	113(37.9)	9(7.8)	0.9(0.2-4.5)
Third & more visit	27(9.1)	2(1.7)	1
Husband's Occupation			
Employed in any organization	77(20.5)	6(3.4)	7.2(2.6-20.3)*
Unemployed	221(79.5)	110(96.6)	1
Husband's educational level			
Formal education	145(49.0)	20(17.2)	4.6(2.7-7.8)*
Informal education	153(51.0)	96(82.8)	1
Husband's Age			
21-30 years old	137(46.0)	44(37.9)	3.1(1.0-9.4)*
31-40 years old	119(39.9)	45(38.8)	2.6(0.9-7.9)
41-50 years old	35(11.7)	20(17.2)	1.75(0.5-5.7)
51-60 years old	7(2.3)	7(6.0)	1
Length of years in relationships			
< 5 years	120(40.3)	43(37.1)	1.7(1.0-2.9)*
5-10 years	105(35.2)	28(24.1)	2.3(1.3-4.0)*
> 10 years	73(24.5)	45(38.8)	1

N.B: *statistically significant (p<0.05) 1= reference group

Table 3: Association of knowledge about HIV/AIDS, MTCT, PMTCT and acceptance of HIV counseling and testing among ANC attendants at East Gojam public health centers ANC clinic, North-West Ethiopia, March – April, 2011

Variables	Acceptance of HCT
-----------	-------------------

	Yes Num (%)	No Num (%)	Crude OR (95% C.I)
HIV/AIDS transmit through MTCT			
Yes	261(87.6)	60(51.7)	6.6(3.9-10.9)*
No	37(12.4)	56(48.3)	1
HIV/AIDS transmit through MTCT during pregnancy			
Yes	131(44.0)	17(14.7)	4.6(2.6-8.0)*
No	167(56.0)	99(85.3)	1
HIV/AIDS transmit through MTCT during delivery			
Yes	81(27.2)	3(2.6)	14.1(4.3-45.5)*
No	217(72.8)	113(97.4)	1
HIV/AIDS transmit through MTCT during postpartum period			
Yes	248(83.2)	46(39.7)	7.5(4.7-12.2)*
No	50(16.8)	70(60.3)	1
A woman is infected with the AIDS virus, is there any way to avoid MTCT			
Yes	260(87.2)	53(45.7)	8.1(4.9-13.4)*
No	38(12.8)	63(54.3)	1
know the existence of intervention which reduce MTCT			
Yes	225(75.5)	24(20.7)	11.8(7.0-19.9)*
No	73(24.5)	92(79.3)	1
Use antiretroviral drug can reduce MTCT			
Yes	169(56.7)	13(11.2)	10.4(5.6-19.3)*
No	129(43.3)	103(88.8)	1
Avoid breast feeding can reduce MTCT	Yes		
Yes	185(62.1)	17(14.7)	9.5(5.4-16.8)*
No	113(37.9)	99(85.3)	1
Safe delivery can reduce MTCT			
Yes	42(14.1)	4(3.4)	4.6(1.6-13.1)*
No	256(85.9)	112(96.6)	1

N.B: *statistically significant (p<0.05) 1= reference group

Table 4: Association of HIV Counseling & Testing Service factors and acceptance of HIV counseling and testing among ANC attendants at East Gojam public health centers ANC clinic, North-West Ethiopia, March – April, 2011

Variables	Acceptance of HCT
-----------	-------------------

	Yes Num (%)	No Num (%)	Crude OR (95% C.I)
Ever heard of HCT			
Yes	272(91.3)	81(69.8)	4.5(2.3-7.3)*
No	26(8.7)	35(30.2)	1
Source of information Mass media			
Yes	186(62.4)	30(25.9)	4.8(2.9-7.7)*
No	112(37.6)	86(74.1)	1
Source of information Health works			
Yes	201(67.4)	50(43.1)	2.7(1.8-4.2)*
No	97(32.6)	66(56.9)	1
Somebody ever told about the benefit of HIV testing			
Yes			
No	256(85.9)	47(40.5)	8.9(5.5-14.7)*
	42(14.1)	69(59.5)	1
HCT is important for pregnant women			
Yes	253(84.9)	51(44.0)	7.2(4.4-11.6)*
No	45(15.1)	65(56.0)	1
Ever had HCT			
Yes	185(62.1)	13(11.2)	12.9(7.0-24.1)*
No	113(37.9)	103(88.8)	1
Disclose test result			
Easy	251(84.2)	25(21.6)	19.4(11.3-33.4)*
Difficult	47(15.8)	91(78.4)	1

N.B: *statistically significant (p<0.05) 1= reference group

Table 5: Association of role of male partners and acceptance of HIV counseling and testing among ANC attendants at East Gojam public health centers ANC clinic, North-West Ethiopia, April, 2011

Variables	Acceptance of HIV counseling & testing
-----------	--

	Yes Num (%)	No Num (%)	Crude OR (95% C.I)
Male partner accompany			
Yes	59(19.8)	6(5.2)	4.5(1.9-10.8)*
No	239(80.2)	110(94.8)	1
Partners coming with them ever before			
Yes	99(33.2)	33(28.4)	1.6 (1.0-2.7)*
No	199(66.8)	83(71.6)	1
Partners approval			
Yes	251(84.2)	58(50.0)	5.3(3.3-8.7)*
No	47(15.8)	58(50.0)	1
Partners invitation			
Yes	213(71.5)	57(49.1)	2.5(1.7-4.0)*
No	85(28.5)	59(50.9)	1
Partners consultation			
Important	271(90.9)	76(65.5)	5.3(3.0-9.2)*
Not Important	27(9.1)	40(34.5)	1
Need partner involvement before decision on acceptance of HCT			
Yes	233(78.2)	35(30.2)	8.3(3.3-13.2)*
No	65(21.8)	81(69.8)	1
Ever discussed about health issues (HCT) with your partner			
Yes	175(58.7)	34(29.3)	3.4(2.2-5.4)*
No	123(41.3)	82(70.7)	1
Test result was found to be positive, would your partner support			
Yes	171(57.3)	12(10.3)	7.2(3.6-14.3)*
No	38(12.8)	59(50.9)	0.3(0.2-0.6)*
I am not sure	89(29.9)	45(38.8)	1

N.B: *statistically significant ($p < 0.05$) 1= reference group

RENEWABLE ENERGY FINANCING: Towards a Financing Mechanism for Overcoming Pre-Commercialization Barriers of Renewable Energy Financing System in Nigeria

Kaisan Muhammad Usman, Aminu Haruna Isa, and Johnson Oluyemi Ojoso

ABSTRACT

Nigeria is exploiting a new way of boosting its energy supply by utilizing all its available renewable energy resources. As a way of tackling the persisting energy crises in the country, renewable energy technology is being introduced into the nation's energy supply mix. However, the utilization of renewable energy technology innovation process is influenced at different stages by technology-push and market-pull forces that need to be managed and balanced by technology innovators, governments, financiers and eventually end-users. The problem today in Nigeria is the pre-commercialization financing gap that plaques new technology utilization developments and the financing mechanism needed to overcome these problems. Such financing gap is as a result of the quality of financing intermediation, debtor information, cost to banks and financing needs of renewable energy technology firms. In solving this problem, we considered three financing aspects of renewable energy technology that target the enhancement of the financing sector performance, such as sources of financing, policies for mitigating financing barriers and financing mechanism. The result shows that the development of renewable energy technology to sustainable use depends on financing mechanism that targets the pre-commercialization financing gap.

KEY WORDS: Renewable energy, Sources of financing, Policies and Financing mechanisms

INTRODUCTION:

The Government of the Federal Republic of Nigeria has taken bold steps and has put together solid plans and various initiatives to ensure a permanent solution to the country's energy crises, by introducing renewable energy (RE) into its energy supply mix. Policies and strategies have continuously been formulated by the government to project its energy demands and supplies for

- *Engr. Kaisan Muhammad Usman is a Scientific Officer, Energy Management, Training and Manpower Development, Energy Commission of Nigeria, Abuja, Nigeria*
E-mail: engineer.kaisan@gmail.com PH: +2348036943404
- *Mr. Aminu Haruna Isa is a Senior Scientific Officer, Energy Planning and Analyses, Energy Commission of Nigeria and a PhD Candidate in University of Abuja, Nigeria*
- *Engr. J. O. Ojoso is the Director energy Planning and Analyses, Energy Commission of Nigeria.*

decades yet to come. The National Renewable

Energy Master Plan has been developed by Energy Commission of Nigeria (ECN); and this recommended that 23% of the national energy supply mix should come from renewable energy technology¹ (RET). There are several renewable energy technology (RET) projects that are being carried out in the country. However, one of the most important parameters needed to achieve the successful implementation of the National Renewable Energy Master Plan is the financing mechanism of the RE projects. Presently, RE is on the agenda and it has gained a strong momentum, despite this, RE projects still face serious constraints hampering their further development and commercialization. Hence, significant improvements in energy efficiency and conservation and a transition to renewable energy (RE) in Nigeria will require huge investments in the national, state and local energy infrastructures over the coming decade². These investments will

need to come from both the public and private sectors, and they will have to take many forms: including financial incentives from government; loans and capital investment from banks, private investors, venture capital funds and communities; as well as new innovative markets that contribute to the benefits of RET. Financing sources will include venture capital, share raised capital, energy bill charges (public benefits or lines charges), financial institutions, community development funds and general tax revenues.

RETs are unique in that they require higher up-front investment than conventional energy sources, while at the same time providing multiple benefits that are not reflected in their cost. Hitherto, Nigeria lags behind other African countries like South Africa and Egypt, in its support for RE investment³. Innovative strategies and policies are therefore needed to increase investment, spread cost over the life cycle, and reflect the multiple benefits of renewable energy. In this regard, Nigeria needs to:

Make a strong political commitment to RE and their many benefits such as economic development, job creation, energy security and reliability;

Implement ecological tax reform in which financial incentives for conventional energy sources such as oil, gas, coal are significantly reduced and diverted into incentives for RE;

Make changes to public support (tax breaks, subsidies, royalty reductions) of conventional energy to divert investments into RE, especially at the local level; and

Provide and training other programs that build national infrastructure to manufacture, assemble, distribute, install, operate and maintain all types of centralized and distributed RE technologies.

Policy implications: experience from developed countries

The European community has already issued directives on energy efficiency and Green Power and is working on one for Green Heat. This will establish a sound long-term investment environment for RE. In the US, state level policies are leading the way in encouraging investment³.

2.1 POLICIES THAT LEVERAGE INVESTMENT IN RENEWABLE ENERGY

Countries such as Spain, Germany, India, Italy and the United Kingdom as well as some States in the US are leading the world in the implementation and manufacturing of renewable energy systems, such as wind turbines and solar systems. These efforts provide clear examples of what can be quickly achieved when the right policy mechanisms are in place. Their leadership and success is based on a set of common factors⁴:

- Very active political commitment to renewable energy;
- Supportive education initiatives for R&D, training and public awareness;
- Strong incentive systems to achieve wide public participation; and
- The implementation of supportive

Policies such as renewable energy standing-offer contracts or FIT, renewable energy obligations and financial incentives.

TABLE 1
SUMMARIZES SOME OF THE POLICIES USED TO MAXIMIZE INVESTMENT.

Country	Index	Policies
Spain	1	Carbon penalties on fossil fuels Potentials of fixed (standing offer) of market –based tariffs for renewable power sources Solar thermal incentives
United States	2	Renewable energy production tax credit State renewable portfolio standards and renewable energy certificate markets

Germany	3	20-year guaranteed feed-in-tariff/standing offer National target for 2010 and 2020
United Kingdom	4	Certificate based renewable obligations
India	5	Regional (state) feed-in-tariffs/standing offer State level renewable obligations
Italy	6	Renewable obligations and green certificates program (with long term value)

In Germany, the renewable energy sector plans to invest 200 billion euros in renewable energy by 2020, which is 15 times the announced investment in new coal power plants⁵. Germany's goal is to have more than 20% of its energy consumption from renewable energies by 2020. The long-term nature of Germany's commitment to a strong target and a variety of ownership models is responsible for these high investment projections. Successful policy approaches for encouraging investment in Germany have included legislation that guarantees access to the power grid and fixed tariffs over 20 years for private investors in renewable energy generation. Germany has also introduced an energy tax to reduce energy consumption, and it provides subsidized loans for investments in energy efficiency.

The California Solar initiative is a comprehensive set of policies designed to leverage investment in 3GW of solar photovoltaic power systems on new buildings. The initiative includes a public-benefits charge to raise public funds that will provide rebates to builders. The remaining costs will be passed on to building owners through mortgages and leases⁵.

R.E. FINANCING GAPS /BARRIERS IN NIGERIA

Experience shows that the cost of RET falls approximately 20% every time accumulated production doubles². The result of gaps/barriers puts Nigerian RE at an economic, regulatory, or institutional disadvantage relative to other forms of energy supply. These gaps/barriers include continuous subsidies for conventional forms of energy, high initial capital costs of RET, imperfect capital markets, and lack of skills or information, also poor market acceptance, technology prejudice, financing risks and uncertainties. Many RETs are commercially available but still in the cost-reduction phase. They also face many non-price market barriers; however, such as lack of consumer understanding and capability, lack of product and service availability, nevertheless, three broad categories of barriers are discussed in this section⁴.

COSTS AND PRICES THAT UNFAIRLY DISTORT R.E. FINANCING IN THE NIGERIAN MARKETS

There has been argument that RE "costs more" than other energy sources, resulting in cost-driven public decisions that avoid RE. In practice, this comparison can be distorted by a variety of factors. For example, the initial capital costs for RET are often higher on a cost-per-unit basis (i.e., \$/kW), but this argument is not based on the total "lifecycle" costs, that is, the initial capital costs; future fuel costs; future operation and maintenance costs; decommissioning costs; and equipment lifetime. Also, part of the distortion for the comparisons goes like: What are fuel costs going to be in the future? How should future costs be discounted to allow comparison with present costs based on expected interest rates? The uncertainties inherent in these questions affect cost comparisons. Existing analytical tools for calculating and comparing costs can discriminate against RE if they do not account for future uncertainties or make unrealistic assumptions⁶.

i. SUBSIDIES FOR COMPETING FUELS

Large public subsidies, both implicit and explicit, are channeled in varying amounts to all forms of energy, which can distort investment cost decisions. Organizations such as the World Bank and International Energy Agency put global annual subsidies for fossil fuels in the range of \$100-200 billion, although such figures are very difficult to estimate Renewable energy markets in developing countries⁶. Statistics released recently

shows that over 1.8 trillion naira is spent by the three tiers of government (federal, state and local) on oil subsidy between 2006 and 2008. Large subsidies for fossil fuels like this is responsible for lowering final energy prices in, putting renewable energy at a competitive disadvantage if it does not enjoy equally large subsidies.

ii. HIGH INITIAL CAPITAL COSTS

Despite lower fuel and operating costs enjoyed by RE on a life-cycle basis, provides less installed capacity per initial dollar invested than conventional energy sources. Thus, RE investments generally require higher amounts of financing for the same capacity as conventional. Depending on the circumstances, capital markets may demand a premium in lending rates for financing RE projects because more capital is being risked up front than in conventional energy projects. Taxes and import duties may also exacerbate the high first-cost considerations relative to conventional energy technologies and fuels⁸.

iii. UNFAVORABLE POWER PRICING RULES

Due to unlevelled playing field in Nigeria and lack of RE Feed-In-Tariff (FIT), RE sources feeding into an electric power grid may not receive full credit for the value of their power. Two factors are at work. First, RE generated on distribution networks near final consumers rather than at centralized generation facilities may not require transmission and distribution (i.e., would displace power coming from a transmission line into a node of a distribution network). But utilities may only pay wholesale rates for the power, as if the generation was located far from final consumers and required transmission and distribution. Thus, the "locational" value of the power is not captured by the producer. Second, renewable energy is often an "intermittent" source whose output level depends on the resource (i.e., wind and sun) and cannot be entirely controlled. Utilities cannot count on the power at any given time and may lower prices for it. Lower prices take two common forms: (i) a zero price for the "capacity value" of the generation (utility only pays for the "energy value"); (ii) an average price paid at peak times (when power is more valuable) which is lower than the value of the power to the utility—even though the renewable energy output may directly correspond

with peak demand times and thus should be valued at peak prices⁶.

3.2 LEGAL AND REGULATORY

i. LACK OF LEGAL FRAMEWORK FOR INDEPENDENT POWER PRODUCERS

Up till in now, many countries' power utilities are in the hold of monopoly control on electricity production and distribution. A capital example is Nigeria. In these circumstances, in the absence of a legal framework, independent power producers may not be able to invest in renewable energy facilities and sell power to the utility or to third parties under so-called "power purchase agreements."⁶

ii. TRANSMISSION ACCESS

Transmission access is necessary because some renewable energy resources like windy sites and biomass fuels may be located far from population centres. This access is also necessary for direct third-party sales between the renewable energy producer and a final consumer. Utilities may not allow favourable transmission access to renewable energy producers, or may charge high prices for transmission access.

iii. LIABILITY INSURANCE REQUIREMENTS

Small power generators (particularly home PV systems feeding into the utility grid under "net metering" provisions) may face excessive requirements for liability insurance. The phenomenon of "islanding," which occurs when a self-generator continues to feed power into the grid when power flow from the central utility source has been interrupted, can result in serious injury or death to utility repair crews. Although proper equipment standards can prevent islanding, liability is still an issue.

3.3 MARKET PERFORMANCE

i. Lack of access to credit. In Nigeria, consumers or project developers may lack access to credit to purchase or invest in renewable energy because of lack of collateral, poor creditworthiness, or distorted capital markets. Likewise, in rural areas of Nigeria, "microcredit" lending for household-scale renewable energy systems does not exist. Available loan terms may be too short relative to the equipment or investment lifetime. In Nigeria, RE power project developers will have difficulty

obtaining bank loans because of uncertainty as to whether utilities will honour long-term power purchase agreements to buy the power⁴.

ii. PERCEIVED TECHNOLOGY PERFORMANCE UNCERTAINTY AND RISK

Lack of familiarity with renewable energy technologies can lead to perceptions of greater technical risk than for conventional energy sources. This is because of the little or no experience with them in a new application or region. These perceptions may increase required rates of return, result in less capital availability, or place more stringent requirements on technology selection and resource assessment. Utilities may be hesitant to develop, acquire, and maintain unfamiliar technologies, or give them proper attention in planning frameworks.

iii. LACK OF TECHNICAL OR COMMERCIAL SKILLS AND INFORMATION

Markets function best when everyone has low-cost access to good information and the requisite skills. But in specific markets, skilled personnel who can install, operate, and maintain renewable energy technologies may not exist in large numbers. Project developers may lack sufficient technical, financial, and business development skills. Consumers, managers, engineers, architects, lenders, or planners may lack information about RET characteristics, economic and financial costs and benefits, geographical resources, operating experience, maintenance requirements, sources of finance, and installation services. The lack of skills and information may increase perceived uncertainties and block decisions.

STRATEGIZING R.E. FINANCING TO COMMERCIALIZATION IN NIGERIA

Today, with the pre-commercialization financing state of RE financing in Nigeria, we need to consider three financing strategies in order to move ahead:

Sources of financing: Sources of financing such as loans, investment capital, environmental markets, international facilities and partnerships; including financial institutions that lend to developers of new facilities like solar, wind farms, or bio-fuel production plants. These institutions also lend to energy users who purchase renewable energy equipment.

GOVERNMENT POLICIES FOR LEVERAGING INVESTMENT IN R. E. SUCH AS MARKET TRANSFORMATION, FINANCIAL SUPPORT

Financial incentives such as production or user tax credits, standing-offer contracts that provide a fixed higher tariff for renewable power and/or direct financial assistance in the form of rebates or free installation, are effectively a public source of financing. Regulation includes removing inefficient and conventional investment options from the market through performance requirements in building codes and equipment standards. Legally binding targets for RE can also be set. Market support policies include certification and training, information and technical assistance to users, market transformation, infrastructural development and other programs that remove investment barriers.

RE financing mechanisms: Financing mechanisms such as micro finance, on-bill payment, leasing/rental, local improvement charges, allow the purchaser or developer to pay back a loan or provide a return to investors at a rate less than or equal to the income or savings achieved. For example, micro-credit schemes used in many developing countries allow buyers of solar-home systems to pay for the system at the same rate as they would have had to pay for kerosene or battery charging. If a building efficiency improvement is financed by a municipality and repaid as a local improvement charge, the cost is associated with the property and not the owner, allowing transfer of costs and benefits from owner to owner.

INNOVATIVE FINANCING MECHANISMS FOR THE PRE-COMMERCIALIZATION FINANCING GAP PROBLEM IN NIGERIA

Innovative approaches are needed at the national and regional level to allow the up-front costs of RE to be spread out over the lifetime of the technology, and to monetize the multiple benefits of RE. There are a number of ways to help spread upfront costs over a period long enough to provide a positive cash flow for renewable energy developers: providing revolving funds for micro-finance and renewable energy technology and service bundling programs into larger investments; offering guarantees to reduce loan risk; and providing long-term purchase agreements for renewable power, heat and fuels⁹.

These innovative market financing mechanisms include:

(i) the issuing of green and white certificates; the issuing of certificate for verified RE production referred to as green certificates and are now being applied to energy savings referred to as white certificates. France and Italy have had a white certificate program in place for a number of years, and several EU countries have set up the Euro White Certificate project¹⁰. The market for green or white certificates includes utilities that are legally required to meet renewable energy or energy efficiency targets as well as individuals, organizations and socially responsible corporations that want, voluntarily, to “green” their energy purchasing. Some companies are now also becoming investors—moving beyond purchasing green energy to also investing in it⁹.

(ii) Use of local charges: a new option for financing long-payback energy efficiency and renewable energy improvement in buildings is the use of local improvement charges. The cost of the renewable energy equipment or building upgrade is financed by the municipality and repaid through the property tax system. By associating the cost with the property and not the owner, the cost and benefits are passed from owner to owner, allowing equipment of a much higher cost to be installed. This option would be ideal for financing a distributed generation system (e.g. cogeneration, solar PV) that sells power to the grid under a long-term standing-offer FIT contract.

(iii) Risk reduction: is another important practice for developers of RE. Delays, cost overruns, resource uncertainty (in quantity and in price), technical risk, maintenance costs, sales price/volume, renewable energy premiums and tax environment are all risks that need to be identified and managed. Because investors and lenders typically have a low threshold for risk, RE investments will only grow as low-risk projects become available.

DISCUSSION OF RESULT AND RECOMMENDATION

The suggested implementation strategies are divided into three for easy implementation by the various stakeholders.

STATES AND LOCAL GOVERNMENTS STRATEGIES

States and local governments should come up with programs that will maximize private, community and public investment in RE. These programmes should be part of a politically-supported, comprehensive policy strategy which includes targets and milestones, financial incentives, new funding sources, regulations, capacity building and training. The following are ways to supporting RE investment in states and local government levels:

- Provide incentive mechanisms such as FIT, renewable portfolio standards (RPS), renewable energy certificates.
- Take action to remove all barriers to renewable energy investment and installation, including requirements for solar readiness in building codes; updates of electrical, plumbing and building codes; and training of inspectors.

Support policies that encourage community investment in and ownership of renewable energy systems including community finance funds, training and legal framework.

FEDERAL GOVERNMENT ENABLING POLICIES AND SUPPORT

The federal government should play a leadership and enabling role to increase investment in and financing of renewable energy as part of a national strategy so that Nigeria can become an attractive place to invest in these resources. Specifically, the objectives of the federal government should be to remove barriers, level the playing field, and maximize private and public investment:

LEADERSHIP

This is the role needed for the federal government to take in implementing the recommendation

FINANCING

- Establish a national RE investment facility with major banks and credit unions. Investment targets should be set for each technology and end-use.
- Encourage the private sector to establish more venture capital funds for investment in/and debt financing of RE.

- Develop and implement a national RE industrial development and infrastructure action plan which will include financial support for commercialization and cost reductions in manufacturing; training and certification of the designers, installers, operators and inspectors of large and distributed systems; financial incentives for manufacturers, builders, suppliers, etc.; and risk-reduction strategies for project developers.
- Support innovative financing strategies such as a national tradable certificates system (green and white certificates) for RE investments that would work with national or states portfolio standards and municipal financing using local improvement charges.
- Make changes to the public support (tax breaks, subsidies, royalty reductions) of conventional energy to divert investments into RE, especially at the local level.
- Provide training and other programs that build a national or state level infrastructure to manufacture, assemble, distribute, install, operate and maintain all types of centralized and distributed RETs.
- Take action to remove all barriers to RE investment and installation, including requirements for solar readiness in building codes; updates of electrical, plumbing and building codes; and training of inspectors.

Support policies that encourage community investment in and ownership of RE systems including community finance funds, training and legal framework.

PRIVATE SECTOR INVESTMENT AND OTHER STAKEHOLDERS

Private and other non-governmental organizations can also play an important role in financing the transition to sustainable RE production and utilization:

- Municipalities can encourage community investment in RE through the establishment of community power corporations, green funds and the use of local improvement charges for project financing.

- RE industries can join with NGOs and other stakeholders in holding finance forums, advocating more support for investment and local manufacturing. They can work with government and stakeholders to increase Nigeria's investment attractiveness.

- NGOs can work with all stakeholders to lobby for investment supportive policies in Nigeria. They can participate in all global networks that support the establishment of an International Renewable Energy Agency and Investment Bank.

Socially responsible corporations and institutions can purchase green energy (power, fuels, and heat) by buying tradable certificates and investing in community power and fuels projects.

International financial institutions can set meaningful and ambitious RE targets, with part of the funding allocated to the development of energy commodities markets in the six-geo political regions of the federation. The granting capacity of the Global Environmental Facility (GEF) should be increased. A renewable energy investment attractiveness index that includes sustainable and local development value in its metrics needs to be developed for all the states.

CONCLUSION

RE financing in Nigeria is crucial to the maximizing of national transition to RE utilization on a larger scale. Today, RE has not gone far beyond the R&D, and to achieve a transition to commercialization stage, we suggested in this paper innovative financing mechanisms such as: the issuing of green and white certificates for verified RE production and carbon reduction respectively; the use of local improvement charges including FIT to pass benefits from utilities to investors; risk reduction for institutional framework, are all needed because of the unique features of RET. Successful implementation of the Renewable Energy Master Plan, Government commitment and policies to leverage investment are the keys to success.

States and local governments can play important roles in attracting investment through standing offers and Renewable Portfolio Standards. The federal government must show leadership at the national level and international level by providing

financial incentives, removing barriers and levelling the playing field.

REFERENCES

ECN (2005). National Renewable Energy master Plan.

Global Village Energy Partners (2010), Innovative financing for low-carbon development in Africa: <http://www.gvep.org/>

International Energy Agency. (2003). Renewable Energy Policies and Measures in IEA Countries. Paris.

Press Release (2006): "Renewable Energy Branch wants to invest 200 billion Euros in Germany." Information Campaign for Renewable Energy, <http://www.unendlich-viel-energie.de/index.php?id=239>

Renewable Energy World (2006), Vol. 9, No. 2, p.22.

Fred Beck and Eric Martinot 2004: Renewable Energy Policies and Barriers Encyclopedia of Energy, Cutler J. Cleveland, ed. (Academic Press/Elsevier Science).

Momodu S. (2010) "Fuel Subsidy: How FG Burnt N2 trillion in Four years." This Day, February 2 and May 30.

Martinot, E., Chaurey, A., Moreira, J., Lew, D., Wamukonya, N. (2002). Renewable energy markets in developing countries. Annual Review of Energy and the Environment 27: 309-348.

Jennifer I., Julie A. and Matthew B. (2008): Appraisal Guide for Microfinance Institutions: Revision of 1999 Appraisal Handbook.

Morrison J. (2006). "Moving Renewable Energy beyond Green Tags." North American Wind Power, April, p. 26.

Change in the Concept of Equilibriums and Reformulation for Newton's Second Law in Presence and in Absence of Gravitational Field

Bishal Banjara

Abstract-Newton's second law which deals with measurement of force that acted on a body by change in velocity upon time taken or rate of change of the linear momentum with time represented mathematically by $F=ma$, is a way to measure force acted on that body in absence of external force or external force exerting field. But this formula can't measure force completely in presence and in absence of external force exerting field such as in Gravitational field, although instead of $F=ma$ there it is replaced by $F=mg$. So, a complete way to measure force acted on a body is of common essence. For the purpose I have intended forward to formulate a complete way to measure force acted on a body in such Gravitational field and in absence of such field. It is denoted by formula $F=mg\sin\theta+mV_0V_f/R$ in presence of Gravitational field and $F=ka$ in absence of such field where, letters have their usual meanings.

Index terms- force measurement, equilibriums, Gravitational field, Newton's laws, inertial frames of reference, terminal velocity, coefficient of viscosity.

Introduction:

Almost all of the experiments on Mechanics till date are based on Newton's laws of motions in terms of force measurement. Particularly Newton's second is standing as the Pillar of Mechanics because it deals from micro to macro level to calculate the net force presence on a particle and on a body with huge mass. Furthermore the concept of Newton's first law is only the part of Newton's second law. Some of the examples implementing these laws are Stoke's law, Millikan oil drop experiment, etc. One thing we can observe in Mechanics is that whenever the Newton's laws of motion deals to illustrate the phenomenon or the nature of balancing force occurs, there will come the concept of Equilibriums to deal the balancing of the forces those get exerted on a body. In above mentioned experiments also the concept of Equilibrium is referred from Newton's First law that states that all inertial frames deal acting no force. It means whenever a body is allowed to move (in absence of external force) in uniform velocity, it will continues its motion permanently i.e. due to inertia of motion (in absence of external force) and on similar way if the same body is allowed to have a rest on a certain point it will too permanently stay at same point forever i.e. due to inertia of rest(in absence of external force). From this kind of view the concept of Equilibrium is divided into two parts. The phase moving with uniform (following inertia of motion) velocity is referred as Dynamic equilibrium and the phase having no motion (following inertia of rest) is referred as Static equilibrium. But the matter is that though these phases achieved by the same body are different but both of these posses zero resulting forces at all. In the experiments mentioned above, a term 'terminal velocity' which allows to fall the oil drops or piece of iron ball in downward direction uniformly from a particular level of medium, is referred as the phase of dynamic equilibrium where it is said that the net down ward force(weight W) is equalized

by the up-thrust (U) and the viscous force(F) that acts on surface of falling body (as like a resistive force) inside the medium. i.e. $W=U+F$.

Body:

Now the matter arise here that, how this phase of gaining uniform velocity can be referred as the phase of Equilibrium or say the phase of balancing forces? According to Newton's first law there must not be any other external force that could come into action to have the state of Equilibrium but here the viscous force which resists the motion of the body is representing as an external force which is taken wide openly. Is it appropriate to refer such phase as a balancing state of forces? And more strongly if we put the body directly at a point where it gained terminal velocity previously, will the body remains stationary at same point where we put it? No, it will not remain at the point we placed it on certain level of medium. It will slowly comes down at the bottom. It means that there must be some quantity of force that drags the body downwards due to Gravitational force exertion to that body by earth. But from the Newton's first law that body must be remained at the same point where we put it obeying inertia of rest. The answer to these questions can't be given by the Newton's laws of motion. What mattered here is that Newton's laws of motion can't be applied in presence of external force and in presence of such external force exerting fields such as in Gravitational field with respect to mass and in Electrostatic field with respect to charge to deal the phase of Equilibrium of forces. If we perform the same experiments in absence of Gravitational field (say space for consideration) then the body or oil drop or the spherical iron ball will not gain any motion though we put it on the surface of medium or liquid there the body does not follow as it follows in Gravitational

field. It means the body will not fall from the medium though if we let it on surface of medium.

So, there must be change in the concept of Equilibriums to deal with the balancing state of forces in presence of external forces and external force exerting fields. The two types of equilibriums say Static and Dynamic have respective regions for application. As Static equilibrium deals with balancing of the forces those just get cancelled out with zero displacement of body, it must be applied in Multi-body system at a phase when the body remains at rest though it has interactions of different forces but the Dynamic equilibrium that deals with balancing of forces with uniform displacement of body it can be possible only when the body move in orbit with orbital velocity though there is change in direction continuously. But such changing of direction will not be there if we trace different perpendicular lines from the different points on the circumference of earth surface. It seems quite odd but in reality moving on orbit is actually due to balancing the value of centripetal and centrifugal forces in against of one another with uniform angular displacement of the body is due to the dynamic equilibrium. But no such concepts of Equilibriums can be applied on Single body system because, a single body neither get interacted with other forces nor will it have effect of its own. Here, let us introduce the terms Single and Multi-body systems.

(1)Single body system: It is the system where only one body exists and the body must not be in influence of other disturbances like Gravitational field, Electrostatic field, etc. Perhaps it may not be possible or say such body system region could not be possible in reality if the entire universe is holding due to Gravitational forces and their interactions with different masses. It may obey Newton's first law of motion, but not the second law. One will surely get puzzled by this statement but no matter I will show here how this could be possible.

Let us take a mass 'm' having its magnitude as 1 kg on near about the surface of earth. Its weight will be 10N on the surface of earth, if we take it away from the surface of earth to the infinity the magnitude of its weight will slowly goes on decreasing and finally at infinity its weight will be zero following Newton's law of Gravitation i.e. $F = GMm/r^2$. Here infinity represents a region what we call as space now where gravity is assumed to be zero. It shows that 1kg mass will have weight 10N approximately on the surface of earth and 0N in the infinity. It means a man carrying that mass 'm' on earth will continuously get suffered by 10N tension and the same man carrying that same mass will get no tension in the infinity. Now if we go on increasing the mass from 1kg to 100kg, the man can't carry the masses as the weights of those masses goes on increasing on the surface of earth but there will be no effect on changing the magnitude of masses in infinity, the man

will feel as if there is no change on masses. He can take all these masses anywhere as his wish. Say that man allows mass 'm' to move with some constant mechanical force (F) acting on that mass 'm'. The mass will move on constant accelerating (a) condition. If the mass 'm' is replaced by another mass of greater magnitude then at this condition, also the mass must move on same acceleration(a) showing that mass is independent to the force applied on a body in space but depends upon the change in velocity with time(a). So, force applied will be directly proportional to the acceleration only. Thus,

$$F \propto a$$

$\therefore F = ka$ on integrating this equation with respect to velocity (not with distance because, distance covered can't be isolated from time that it required), we get

$$\int F \cdot dv = \int k/t(V_1 - V_2) \quad (1)$$

$$\therefore Power(P) = k/2t(V^2)$$

Where, k is the value that will be universal constant, if we take unit of force as Newton (N) = kgm/sec^2 . This k will be independent on the magnitude of mass bearing the force in absence of Gravitational pulling force. And V_1 and V_2 are the final and initial velocities of the mass 'm' within time interval 't'.

(2)Multi-body system: It is the system where two or more than two bodies co-exist and these bodies make their influences of interactions effective with one another. Here influences of interactions means the effects of Gravitational fields, Electrostatic fields, etc. It is the real system where we are existing. Our universe, galaxies, solar system, etc. are good examples. It's the system where Dynamic equilibrium and Static equilibrium holds at respective phase of motions. From the Newton's law of Gravitation all the masses in the entire Universe are guided by the Gravitational forces those setup immediately in presence of Gravitational fields. And all the lower masses are revolving around the bigger masses such as satellites are revolving around the planets, planets are revolving around the sun and sun is following the same process around the Galactic centre. If each and every massive bodies are rotating with respect to their mass and Gravitational pulling capacity is true, it can't be easy to find the region where such pulling or Gravitational forces are ineffective. It means every region of space is under the influence of Gravitational fields of certain massive bodies indicating that there is no region in space where the inertial frames of references do really exist. Let's have another such kind of illustration that default the application of Newton's laws of motions focusing basically on his second law that states that time rate(dt) of change of

linear momentum (dp) is directly proportional to force applied (F) and takes place in the direction of force applied, represented mathematically by $F = dp/dt = m dv/dt = ma$. Thus, $F = ma$ is the mathematical equation that is used in order to measure the quantity of force that a body possesses or a body is applied with, taking no reference to Gravity. And $F = mg$ is also a similar equation used in order to measure the quantity of force taking reference to Gravity. But this equation is not also a complete format equation that can measure the net force acted on the body.

Let a body with mass 'm' is thrown with certain velocity 'v' in absence of Gravitational field, it follows the inertial frame of reference, tracing the linear path. If the same mass 'm' is projected in horizontal form in presence of Gravitational field say taking earth as source for Gravitational field with same velocity 'v' then it will not follow the inertial frames of reference, tracing the curve path depending upon the limit of velocity we have thrown. If the limit of velocity is well enough to meet at the level of orbital velocity then the mass 'm' will continuously move on uniform velocity tracing the circular path that introduces the maintenance of centripetal and centrifugal forces. If the velocity is greater than the orbital velocity the path traced will be parabolic around the earth and if the velocity is lesser than orbital velocity the path traced by the mass 'm' will also be parabolic towards earth. One thing we need to take care is that below the orbital velocity, a reducible quantity due gravity (g) reduces the direction of the body towards the earth's centre so the path traced will be parabolic. Similarly, above the orbital velocity body tends to escape due to increasable quantity so the path traced will be either parabolic or hyperbolic. Here the terms reducible and increasable quantities are respective components of gravity (g) that will be notify later on. It is the case whenever the body is thrown at horizontal form. If the body is projected in vertical form, it will reach at certain height and come to same point again. But if we project or let to move the body in any direction then there the body will itself gain acceleration or we need to apply force to gain acceleration to that body. With reference to this kind of phenomenon I have formulated a formula that could correctly measure the net force presence on the body in a sense of force applied on that body or force gained by that body. We must consider that the horizontal line passing along the surface of earth as the reference line that separates the condition of force applied or force gained.

The parts above this horizontal line represent the sense of force applied and the parts below this line represents the sense of force gained. But this line itself represents the sense of force applied. This horizontal line is nothing other than the line that passes tangentially through the earth surface on each and every points of its circumference. Now let's calculate the force in any direction:

In terms of force applied:

As the horizontal line itself represent the sense of force applied, at which all the masses that come to orbit the earth must lie or must trace the path along this line, they must have an orbital velocity to remain along this line such that the angle between them remains zero degree. It refers the body to be in equilibrium maintaining the centripetal and centrifugal forces in balance to one on another. It is the condition of Dynamic equilibrium where these centripetal and centrifugal forces are balanced to each other with uniform angular displacement of mass 'm'. At this condition weight of mass 'm' (mg) is balanced by force (mv_0^2/R) i.e.

$$mg = mv_0^2 / R$$

$$\therefore a_1 = g = v_0^2 / R \quad (2)$$

Where, v_0 is the orbital velocity and R is the radius of earth. Now let us suppose that the mass 'm' be projected or moving above this horizontal line making certain angle to this line say angle θ . At this time mass 'm' will now either form projectile motion or it will escape from the surface of earth depending upon the limit of velocity. It is a complex state which is not defined by present science in terms of force measurement completely though it has been illustrated under the equation of motion in projectile chapters. Here as soon as the mass 'm' is projected or allow to move, there comes the role of Gravitational force in terms of gravity (g) to affect the motion of that projectile. At this condition factor affecting the acceleration due to gravity (g) is only the distance travelled (h) from a fixed point that correspondingly even affect the height travelled (p) as shown in figure below:

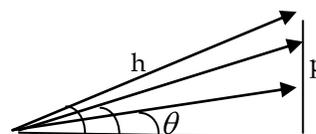


Figure 1: showing the effect of gravity to mass 'm' after projection.

As the body tends to gain height (p) with increasing order of θ then acceleration due to gravity stands in opposition i.e. θ increases causing decrease in the distance travelled (h) by mass 'm' for a constant force. But a fixed gravitational mass (here earth) has fixed value of gravity (g) within a certain region of its gravitational field. So in this case 'g' will be resolved accordingly with $\sin \theta$. Therefore,

acceleration that needs to share with 'g' by mass 'm' can be taken as:

$$a_2 = g \sin \theta \quad (3)$$

Here, for the range of $\sin \theta$ at $(0-180)^\circ$ its value will be positive and positively maximum at $\sin 90^\circ$ so we need to apply more force in terms of gaining acceleration at this case. But up to present this g is taken 'g' only without componential division i.e. without resolving with sine factor. Thus, if v be the instant velocity of body, net acceleration acting on mass 'm' in the force applied direction will be

$$a = a_1 + a_2 = v^2/R + g \sin \theta \quad (4)$$

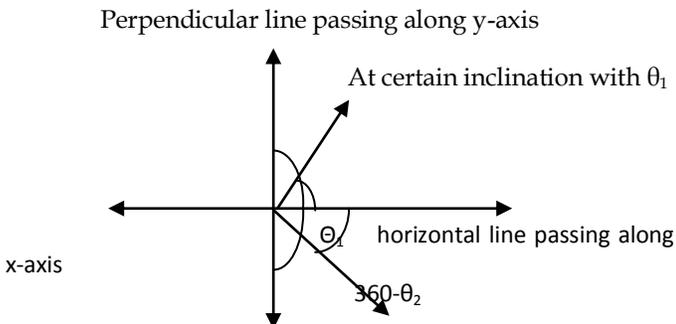
In terms of force gained:

Whenever the mass 'm' moves in downward direction we need not need to apply force but gains acceleration itself where range of θ limits within $(180-360)^\circ$ whose value will be negative and negatively maximum at $\theta=270^\circ$. At this condition the body will be under free fall equalizing the value of 'g' with ' v^2/R '. In this condition the numerical value doesn't get equalize to each other if mass 'm' is allowed to fall from certain height, let β be any numerical value kept on RHS that equalizes RHS and LHS. So the exact equation will be

$$g = \beta v^2/R \quad (5)$$

This particular case will be obtained only in absence of air resistance. In presence of air or any medium that obstacles the motion of the body, $g > \beta v^2/R$. And the difference between g and $\beta v^2/R$ measures the net anti force caused due to the medium. It is called as up-thrust. This equation (5) is exactly the same form as it is in above equation (4). Only net change in acceleration ' $a'=0$ ' in absence of air resistance. But the form is same if there is any disturbance. If the value of $g < \beta v^2/R$, we need to analyze out that an extra source is applying force to the body in same direction of motion. Such conditions will be deal in more detail at below for calculating the value of β .

From these above two cases we can generally represent out a figure that can illustrate the meaning more clearly.



At certain declination with θ_2

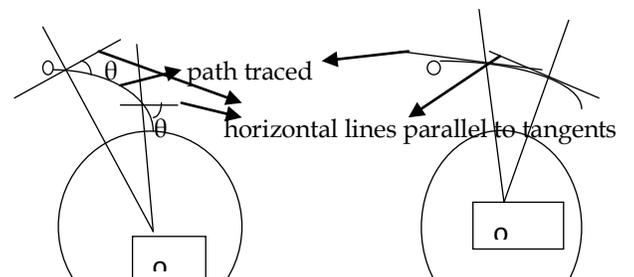
Figure 2: showing the componential division of 'g' with different angles at different positions

On similar way this kind of phenomenon can more clearly be shown by the different motions of mass 'm' in different way with or without earth as a source for Gravitational pull as like in figure below:

	Thrown	mass 'm'	mass 'm'
Thrown			
Mass 'm'			
Path trace is			
Linear in			
Absence of			
Gravitational	for $v < v_0$, path	for $v = v_0$, path	for $v < v_0$, path
Field of earth	traced will be	traced will be	traced will be
	Parabolic around	circular around	parabolic
	the earth	the earth	around earth

Figure 3: Showing the path followed by mass 'm' in presence or absence of Gravitational field thrown with different velocities

In this above figure, the paths traced by mass 'm' is different with vary in limit of velocities, these paths continuously deviate their directions in presence of Gravitational field. Here, we can use the equation (6) to measure the net acceleration acting on mass 'm' more correctly. For this purpose we need to calculate the net change in direction i.e. θ in every points of the path it follows. What geometrically shows that the circular and parabolic paths (near orbital velocity and below it) will have no change in θ with respect to each horizontal lines those passes tangentially at every points on the surface of earth but the parabolic path around the earth will have continuous change in θ . For our experimental purposes we need to calculate θ to deal for such condition. Let's deal with parabolic and circular paths with their illustrations below in figure:



passing from the surface of earth

Figure (a)

figure(b)

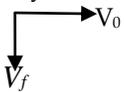
Figure 4: Showing that there will be no change in θ for both

Parabolic (fig.a) and circular(fig.b) paths traced by mass 'm' on moving.

These are the cases when the mass 'm' is projected with different velocities. This shows that below the orbital velocity and below or near about $\theta=0^\circ$, if any mass is projected then there will be always the conservation of kinetic acceleration ($\beta v^2/R$) and potential acceleration ($g \sin \theta$). But when there is presence of external source that provides the force to the mass 'm' then such conservation will be break and force calculation can be done by using the equation(4) for any kind of motion on earth surface or say on any Gravitational surfaces.

Calculation of value of β :

Suppose, mass 'm' is allowed to fall from near to the infinity towards the centre of earth. On falling downwards, velocity goes on increasing order as the distance gets shorter and shorter from the centre of earth. On falling downward, what we need to note is that as the body falls down at any particular point of falling, the velocity measurement gives the velocity for orbital velocity i.e. the tangential velocity required to orbit the earth surface will equalize this particular point velocity. It will be more clear from the figure below where, v_0 is the orbital velocity required at a particular point to gain kinetic acceleration v_0^2/R and V_f is the falling velocity acquired by mass 'm' at the same point. And this V_f is taken as the instant velocity in any direction for the final procedures.



Showing that $V_0=V_f$

But if the mass 'm' is allowed to fall from certain height, we can't have $V_0=V_f$ at any particular point of falling. It means for the body falling from the infinity that will possess the velocity limit in the form of orbital velocity else will possess less than that at any particular point of falling. This shows that for the body falling from infinity will have value of $\beta=1$ and the one falling from any height will have β other than 1. One thing we need to consider out is that a body which may be either in the rest or in the motion at any particular region of gravity, will have the tendency to be get

equalized to the value of gravitational mass at that region. Thus, β must be greater than 1 for falling from general height. And same value of β will be there that neglects the direction of motion at particular level of gravity because, θ recovers the direction of velocity. But in terms of magnitude β depends upon the velocity of the body. It means β depends upon the value of instant velocity in comparative to the value of orbital velocity. So, value of β suits to the ratio of V_0 to the V_f i.e. $\beta=V_0/V_f$ So, the actual force measuring formula will be

$$F = (V_0/V_f)mV_0^2/R + mg \sin \theta$$

$$\text{or, } F = m(V_0V_f)/R + mg \sin \theta \quad (6)$$

Again on integrating equation (6) with respect to velocity, we will get

$$\int F.dv = \int m\{(V_0V_f)/R + g \sin \theta\}.dv \quad (7)$$

$$\therefore P = mV_f\{V_0V_f/2R + g \sin \theta\}.$$

Conclusion 1:

Points to be noted from the equations (1), (6) and (7) are majorly listed below:

- a) If the body gains no velocity at a particular region of gravity but remains stationary at that case V_0V_f/R will be 0, but $a=g$ indicating that force at that time gives weight of the body. It is the real condition of static equilibrium.
- b) A body either in rest or a moving in any direction will always have a tendency to remain with the force that gets exerted by Gravitational pull or simply weight of that body.
- c) There is no need to be change in momentum to have a force magnitude on a body. So, a constant momentum will have constant force. It doesn't link up with time factor. It means a body resting at any point will remain at that point forever with a magnitude of force as its weight up to infinite time and similarly the body orbiting the Gravitational mass will remain at position forever up to infinite time.
- d) Power is the ultimate step that we can obtain or calculate, not the energy which avoids time link to cover a distance. And mass is the factor that link up with force only in presence of Gravitational pull but is independent in absence of such pulling.

Thus, in general sense, the net force gained by body or applied on a body is always comparative to the value of gravity 'g' of any Gravitational masses on its surface. And the force gained by the body in any direction is calculated as $F = m(V_0V_f)/R + mg \sin \theta$ where terms have their own usual meanings. Also, there is no need of motion to have a force magnitude on a body.

One more illustration that needs to be change from present

concept which is dealing with classical concept of Newton's laws of motion:

Millikan's oil drop experiment:

In Millikan's oil drop experiment in first case there the oil drops are allowed to fall with terminal velocity and this phase is referred as the phase of Dynamic Equilibrium saying that the weight (w) of each falling oil drops is balanced by the air particles with up-thrust (u) as there is no change in velocity to have a force magnitude on oil drops from Newton's second law. And on falling each oil drops suffer a constant resistive force that opposes the motion of oil drop naming this force as Viscous force(F₁). Thus the weight(w) of each oil drop is balanced by sum of these up-thrust(u) and viscous force(F₁). i.e. Weight(w)=viscous force(F₁)+up-thrust(u)

If σ be the density of air, V₁ be the velocity acquired by the oil drop, r be the radius of oil drop, η be the coefficient of viscosity of the medium and ρ be the density of oil drop then the actual equation to balance the weight of the oil drop is given by $Mg=6\pi\eta rV_1+4/3\pi r^3\sigma g$ (8)

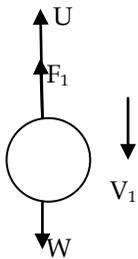


Figure 5: showing that oil drop is falling under terminal velocity V₁

But from the equation (6) above net force that opposes the motion of oil drop will be

$$F = m(V_0V_f) / R + mg \sin \theta$$

$$or, F = m(V_0V_f) / R - mg \tag{9}$$

here $\sin \theta = -1$, m is the mass of the oil drop and $V_f = V_1$. In equation (8) the meaning of up-thrust (u) and in (6) the meaning of anti-force(F) is same. So we can arrive at a relation as viscous force and the kinetic force is same. so,

$$6\pi\eta rV_1 = mV_0V_f / R$$

$$or, \eta = 2/9(V_0 / R)r^2\rho \tag{10}$$

Similarly, in second case when such oil drops are allowed to move in upward direction with applying the force (Fe) by electric field in V₂ terminal velocity acting F₂ as

viscous force. Then the viscous force will change the direction such as shown in figure below:

$$Fe + U = F_2 + W$$

$$or, Fe = 6\pi\eta rV_2 + mg - 4/3\pi r^3\sigma g$$

$$or, QE = 6\pi\eta rV_2 + 4/3\pi r^3(\rho - \sigma)g \tag{11}$$

where, Q is the amount of charge that oil drop carry and E is the electric field intensity applied to the oil drop.

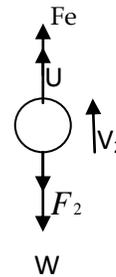


Figure 6: showing that oil drop is moving upward with terminal velocity V₂

Similarly from the equation (6) above net force acting on the oil drop will be $F = m(V_0V_f) / R + mg \sin \theta$ but here Fe is the force applied on oil drop that will move the oil just opposite than that of previous, Fe and U in combination with each other tend oil drop to move in upward direction. So F will be shared by both Fe and U. Therefore,

$$F = m(V_0V_f) / R + mg \sin \theta$$

$$or, Fe + U = mg + mV_0V_2 / R$$

$$or, QE = 4/3\pi r^3(\rho g + \rho V_0V_2 / R - \sigma g) \tag{12}$$

On equating the values of QE from the equations (11) and (12) gives the relation

$$\eta = 2/9(V_0 / R)r^2\rho \tag{13}$$

Conclusion 2:

Thus, in equations (10) and (13) all values remains the same for same volume of oil drops. So, $\eta \propto \rho$ only, predicting out that η is the property of the oil drop(falling body) not of the air(medium that allows the body to fall). This shows that viscous force can't be defined as motion opposing force. It links to the medium just for terminal velocity not for value

of η , η totally depends upon the nature of body which is allowed to fall. This is the reason for the question, why all bodies of varying densities don't fall with same terminal velocity in the same medium. And thus the value will be different for different bodies with vary in their densities in the same medium though they have equal volumes.

References:

[1] "Sir Isaac Newton: The Universal law of Gravitation." Astronomy 161 The Solar System Retrieved 2008-01-04

[2] "Britannica online Encyclopedia"

[3] D.S. MATHUR, "Mechanics"(2nd edition) S.CHAND AND COMPANY LTD.

[4] S.K. GUPTA, J.M. PRADHAN, "A Text Book of PHYSICS" Part-I, SURYA PUBLICATIONS

[5] Pitri Bhakta Adhikari, Bhoj Raj Gautam, Lekha Nath Adhikari, "A Text book of PHYSICS" VOL-I,(4th revised edition)

[6] Manu Kumar Khatri, Manoj Kumar Thapa, Bhesha Raj Adhikari, Arjun Kumar Gautam, Parashu Ram Poudel, "Principles of Physics"

[7] "Physic Forum" discussion

[8] "Answers.com"

[9] "WIKIPEDIA" searching

Dropwise and filmwise condensation

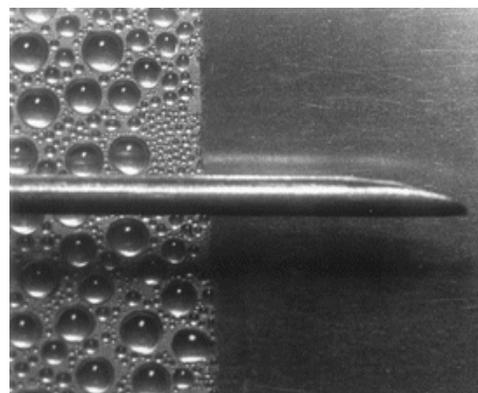
Saurabh pandey

Abstract—The paper reviews progress in dropwise condensation research from 1930 to the present. Particular attention is given to heat transfer measurements, theory, effect the presence of air in the condenser has on the heat flux and surface heat transfer coefficient.. This experiment would be used in any industry which is trying to increase the efficiency of heat transfer. Subsequently, more accurate measurements have shown good consistency and the theory of the dropwise and filmwise condensation have become better understood.. The balance of evidence suggests that dropwise condensation is a more effective method of heat transfer than filmwise condensation, and the presence of air instead vapour significantly reduces the heat transfer..The paper contains the experimental steps involve in performing the experiment precautions and final result of the experiment. The exact reading and values of heat transfer coefficient in both type of condensation is still not fully understood.

Index Terms—Condenser, Dimension less numbers, Dropwise condensation, Filmwise condensation, Heat transfer coefficient, Water flow rate, Mass flow rate, Non condensing gases

1 INTRODUCTION

Proceedings Steam may be condense onto the surface in two distinct modes, known a "filmwise" & " dropwise" For the same temperature difference between the steam & the surface, dropwise condensation is much more effective then filmwise condensation & for this reason the former is desirable although in practical plants it rarely occurs for prolonged periods. In filmwise condensation a laminar film of vapour is created upon a surface. This film can then flow downwards, increasing in thickness as additional vapour is picked up along the way .In dropwise condensation vapour droplets form at an acute angle to a surface .These droplets then flow downwards ,accumulating static droplets below them along the way. The second objective of this is to investigate the difference in heat flux between the two forms of condensation for the same set of conditions. Third objective is to investigate what effect the presence of air in the condenser has on the heat flux and surface heat transfer coefficient. This experiment would be used in any industry which is trying to increase the efficiency of heat transfer. An example of this is any vapour power cycle such as the Rankine cycle. By increasing the efficiency of the condenser, its operational pressure can be reduced and the overall efficiency of the cycle can be increased.



Dropwise and Filmwise condensation

2 CONDANTION

2.1 Dropwise condensation

By specially treating the condensing surface the contact angle can be changed & the surface become 'non - wettable' .As the steam condenses ,a large number of generally spherical beads cover the surface. As the condensation proceeds ,the bead become larger, coalesce, and then strike downwards over the surface. The moving bead gathers all the static bead along its downward in its trail. The 'bear' surface offers very little resistance to the transfer of heat and very high heat fluxes therefore possible.Unfortunately, due to the nature of the material used in the construction of condensing heat exchangers, filmwise condensation is normal .(Although many bare metal surfaces are 'non-wettable' this not is true of the oxide film which quickly covers the bare material).

2.3 Filmwise condensation

Unless specially treated, most materials are wettable as condensation occurs a film condensate spreads over the surface.The thickness of the film depends upon a numbers of factors, e.g. the rate of condensation ,the viscosity of the condensate and whether the surface is horizontal or vertical, etc.Fresh vapour condenses on to the outside of the film & heat is transfered by conduction through the film to the metal surface beneath. As the film thickness it flows downward & drips from the low points leaving the film intact & at an equilibrium thickness.The film of liquid is barrier to transfer of the heat and its resistance accounts for most of the difference between the effectiveness of filmwise and dropwise condensation.

2.4 Methods and Procedure

Fill up the 5 litre distilled water in main unite by opening the valve.After filling the water close the valve. Start water flow through one of the condensers which is to be tested and note

down water flow rate in rotameter. Ensure that during measurement, water is flowing only through the condenser under test and second valve is closed.

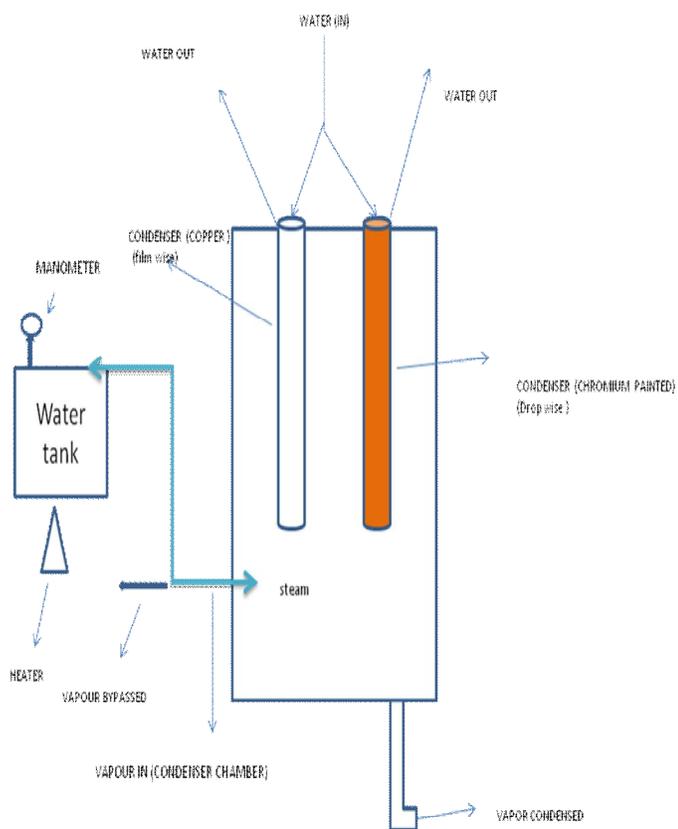
Connect supply socket to mains and switch ON the heater switch

$$1/u = 1/h_i + 1/h_o \times D/D_o \text{ kcal/hr-m}^2\text{-}^\circ\text{C}$$

Same procedure can be repeated for other condenser. Except for some exceptional cases overall heat transfer coefficient for dropwise condensation will be higher than that of filmwise condensation.

Result may vary from theory in some degree due to unavoidable heat losses.

The water level should be up to ¾ th of container. Do not start the heat supply unless water is filled in test tube unit. Operate gently the sector switch of temperature indicator to read various temperature. Slowly generation will be start in stream generation of the unite and the stream rises to the test section, gets condensed on the tubes and fall down in the cylindrical space. Record the the temperature of painted condenser ,plane condenser ,water inlet to condenser and water out let to condenser. Depending upon the condenser dropwise and filmwise condensation can be visualized, If the water flow rate is low the steam pressure will rise in cylindrical region and pressure gauge will read the pressure. If the water flow rate is matched then condensation will occur at more or less atmospheric or upto 1 kg/cm² pressure. Observation like water flow temperature of painted condenser ,plane condenser ,water inlet to condenser and water out let to condenser. Depending upon the condenser dropwise and filmwise condensation can be visualized rates, pressure and noted down in the observation table at the end of each set.



Steam pressure kg/cm ²		0.65	0.60	0.50
Water flow rate LPH		156	54.6	14.1
Condenser under test		Filmwise	Filmwise	Dropwise
Painted condenser outer surface	T ₁	79.3	91.00	90.60
Plane condenser outer surface	T ₂	78.3	88.00	94.3
Steam	T ₃	125	124.1	124.4
Water inlet to condenser	T ₄	30	28.3	28.6
Water outlet plane condenser	T ₅	32.7	32.2	-
Water outlet painted condenser	T ₆	-	-	33.4
Ambient	T ₇	32	32	32

- Saurabh pandey is currently pursuing bachelor degree program in chemical engineering in National Institute of Technology, Jalandhar, India, PH-09872833557. E-mail: saurabh.sleeping@gmail.com

Analysis Normally steam will not be pressurized. But if the pressure gauge reads some pressure then properties of steam should

be taken at that pressure or otherwise atmospheric pressure will be taken.

$$^3).(175 \times 10^{-3}) = 0.010445 \text{ m}^2$$

We will first find the heat transfer coefficient inside the condenser under test. For this properties of water are taken at the bulk mean temperature of water e.g. $(T_{wi} + T_{wo})/2$ where T_{wi} and T_{wo} are water inlet & outlet temperature.

Following properties are required:

- ρ_1 = density of water (kg/m³)
- P_r = Prandtl number

Now calculation for Reynold's number

$$Re_d = 4 m / (\pi \times \rho_1 \times \mu_1 \times D_1)$$

Where

- D_i = Inner diameter of condenser. = 1.75 cms
- If this value of $Re_d = 2100$ then flow will be turbulent in pipe.

Now Nusselt number.

$$Nu_l = 0.023(Re_d)^{0.8} \cdot (P_r)^{0.4}$$

And $h_i = Nu_l \cdot k / L$ kcal/hr.m² (w/m²)

Now calculate the heat transfer coefficient on the outer surface of the condenser (h_o). For this properties of water taken at bulk mean temperature of condenser e.g.

$$(T_s + T_w) / 2$$

- Where T_s = Temperature of steam
- T_w = Temperature of condenser wall

Properties needed are

K_2 = Thermal conductivity kcal/ hr-m °C (w/m- °C)

ρ = density of water (kg/m³)

μ = Viscosity of condensate kgf-sec/m² (kg/m.s)

h_{fg} = Heat of evaporation kcal/kg. (540 Kcal/kg)

$$h_o = \{ (h_{fg} \times g \times \rho^2 \times k_2^3) / ((T_s - T_w) \times \mu \times L) \}^{0.25}$$

- where g = acceleration due to gravity = 9.8 m/sec² = 1.27 × 10⁸ m/hr²
- L = Length of condenser = 160 mm

From this value overall heat transfer coefficient (U) can be calculated in

3 SAMPLE CALCULATION

AREA

- Outer diameter of heat transfer surface, d = 19 mm
- Length of heat transfer surface, L = 175 mm
- Heat transfer area $\pi \cdot d \cdot L$ = $\pi \cdot (19 \times 10$

Temperatures:

All temperatures are taken in

- Steam temperature, T_∞ = 124.1
- Condenser surface temperature, =
- Copper condenser, T_{cu} = 88.4
- Chromium temperature, T_{cr} = 91.4

Flow Rate:

- Flow rate = 455 ml/30sec = 0.0546 m³/hr
- Rate of steam condensed = 780 ml/hr = 3.12 kg/hr
- At temperature = $(T_s + T_{steam})/2$ = $(124.1 + 88.4)/2$ = 106.25 = 0.5 kg/cm²
- Pressure gauge =
- Density of steam condensed at Temperature $(T_{in} + T_{out})/2$ = 995.6 Kg/m³
- Now parameters used,

- H_{fg} = 533.3 kcal/kg
- P = 954.3 kg/m³
- μ = 274.4×10^{-6} N-s/m²
- P_r = 1.75
- K = 0.5860 kcal/h-m- = 276.3×10^{-6} N-s/m²
- Heat flux = $U (T_\infty - T_s)$
- Q = $UA (T_\infty - T_s)$

Over all heat transfer coefficient = $m_s \times h_{fg} / (A \times \Delta T)$

$$U = 3.12 \times 533.3 / ((0.010445 \times (124.1 - 88.4))) = 4462.2 \text{ Kcal/h-m}$$

For filmwise condensation

$$h_o = 0.943 \{ (H_{FG} \times \rho^2 \times g \times k^3) / ((T_\infty - T_s) \times \mu \times L) \}^{0.25}$$

$$h = 0.943 \{ (553.3 \times (954.3)^2 \times 9.81 \times (0.586)^3) / ((124.1 - 88.4) \times 276.3 \times 10^{-6} \times 0.175) \}^{0.25}$$

$$h_o = 6304.209 \text{ kcal/h-m}$$

Calculation For h_i :

$$Re_d = 4 \times m_w / (\pi \times D_i \times \rho_1 \times \mu_1) = v \times \rho \times D / \mu$$

m_w = 54.3 kg/hr

volume flow rate

Q = 0.0546 m³/hr

velocity Q/A = $0.0546 / (3600 \times 0.000227)$ m/s = 0.0668 m/s

$$Re_d = 0.0668 \times 994.5 \times 0.017 / (8.01 \times 10^{-6})$$

$$= 1410605$$

$Re_d > 2100$

$$Nu = 0.023(Re_d)^{0.8} \times (Pr)^{0.4}$$

Here, $Pr = 5.204$ at $T_{avg} = (29 + 34)/2 = 31.5$

$$Nu = 0.023(1410605)^{0.8} \times (5.204)^{0.4}$$

$$= 3696.469$$

$$h_i = Nu \times k / L$$

$$= 3696.469 \times 0.609 / 1.6$$

$$= 1360.76117$$

$$T_{steam} = 125$$

$$T_s = 78.3$$

$$T_{avg} = (125 + 78.3) / 2$$

$$= 101.68$$

$$P_{gauge} = 0.6 \text{ kg/cm}^3$$

$$\text{Density} = 958.4 \text{ kg/m}^3$$

$$H_{fg} = 533.4 \text{ kcal/kg}$$

$$K = 0.5883 \text{ kcal/h-m-}$$

$$Pr = 1.75$$

$$Ho = 0.943 \{ (h_{fg} \times \rho^2 \times g \times k^3) / ((T_{\infty} - T_s) \times \mu \times L) \}^{0.25}$$

$$= 0.943 \{ (533.4 \times (958.4)^2 \times 9.81 \times (0.5883)^3 / ((125 - 78.3) \times 282.4 \times 10^{-6}) \}$$

$$= 5885.32$$

Over all heat transfer coefficient

$$Q = m_s \times h_{fg}$$

$$Q = 4.77 \text{ kg/hr}$$

$$Q = UA(T_{\infty} - T_s)$$

$$U = m_s \times h_{fg} / A(T_{\infty} - T_s)$$

$$= 4.77 \times 533.4 / (128 - 78.3) \times 0.010445$$

$$= 5216 \text{ kcal/h-m-}$$

CALCULATION FOR DROPWISE CONDENSATION

Dropwise condensation

Heat transfer area = 0.010445 m²

$$T_{steam} = 118.3$$

$$T_{chromum} = 83.2$$

$$T_{w in} = 28.6$$

$$T_{w out} = 36.5$$

Pressure gauge 0.6kg/cm²
Mass of steam condensed = m_s
Mass of cooled water = m_w
Mass flow rate of cooling water = 220 kg/hr

Steam condensation = 4.27 kg/hr
Where U is heat transfer coefficient

$$Q = UA\Delta T$$

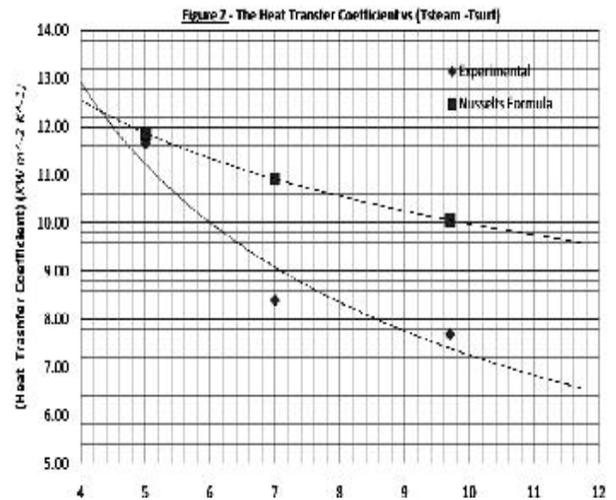
$$U = Q/A\Delta T$$

$$U = 4.27 \times 533.5 / (0.010445 \times (118.3 - 83.2))$$

$$U = 6213.65 \text{ Kcal/hr}$$

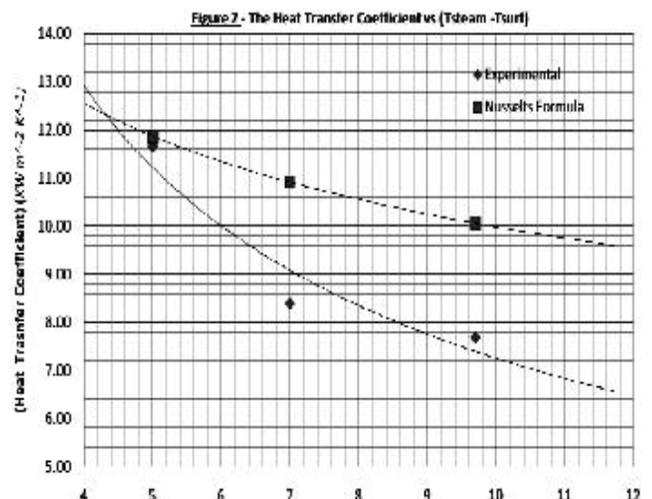
4 CONCLUSION

The final observation is confirmed in the Handbook of Phase Change (2) which quotes that at atmospheric pressure, the Heat Flux in dropwise condensation can be more than filmwise. This can be explained in terms of how the condensation forms on the



condenser.

The vapour drops in dropwise condensation are discrete and are continually formed and released which means that the surface of the condenser is also continually exposed. In comparison, the film created in filmwise condensation always covers the surface of the



condenser (3). As a relatively poor conductor of heat, this film creates a thermal resistance which is the reason why the value for Heat Flux is lower for filmwise in comparison to dropwise condensation (3)

To check the accuracy of the experiment, the values for the Heat Transfer Coefficient in the filmwise condenser were compared to the values which are obtained theoretically using the Nusselt equation (3). Figure 2 shows that the results derived experimentally were of a lower value than of those derived theoretically.

One explanation for this is the presence of non-condensable gases in the steam vapour (1). It shows that for a certain temperature difference, the Heat Flux for a condenser using steam mixed with 5% of air is significantly smaller than pure steam, and the magnitude of this difference increases with temperature difference. In the case of Heat Transfer Coefficients, the value for both steam and steam with air approaches zero, but when the steam is mixed with air it is consistently low.

5 ACKNOWLEDGMENT

With heart filled gratitude I would like to thank Prof. P.K.Mishra (IIT BHU), Prof. K.K.Singh (IIT BHU) for their untiring support and faith they placed in me during the tenure of my work. Moreover, it is my duty to thank my Professors at NIT Jalandhar, namely Dr. A.K.Bansal (HOD Chemical Engineering), Prof. M.K.Jha and all others who encouraged me to pursue this research and even take help from outside institutions for the same. Not to forget, I also intend on thanking the lab assistants at IIT BHU, who were always supporting and helping even at odd hours of my research. Finally, I would like to thank my parents and friends who had me motivated and spirited even when things were on a downfall.

6 REFERENCES

- 1.) Mayhew, Y, Rogers, G. (1992). Engineering thermodynamics: Work & Heat Transfer 4th ed. Prentice Hall.(2)
- 2.) Rose, J, Utaka, Y, Tanasawa, I. (1999). Handbook of Phase Change: Dropwise Condensation. Taylor & Francis.(3)
- 3.) Incropera, F, DeWitt, D. (1996). Fundamentals of heat and mass transfer. 4th ed. Wiley.
- 4.) CHEMICAL ENGINEERING'S HAND BOOK -ROBERT H. PERRY
- 5.) W. Rose, Dropwise Condensation Theory and Experiment: A Review, Proceedings of the Institution of Mechanical Engineers, vol. 216, pp. 115–128, 2002.
- 6.) J.W. Rose, Condensation Heat Transfer Fundamentals, Transactions of the IChemE, vol.76(A), pp. 143–152, 199

Fuzzy Logic Based Gray Image Extraction and Segmentation

Koushik Mondal, Paramartha Dutta, Siddhartha Bhattacharyya

Abstract: Image segmentation and subsequent extraction from a noise-affected background, has all along remained a challenging task in the field of image processing. There are various methods reported in the literature to this effect. These methods include various Artificial Neural Network (ANN) models (primarily supervised in nature), Genetic Algorithm (GA) based techniques, intensity histogram based methods etc. Providing an extraction solution working in unsupervised mode happens to be even more interesting a problem. Fuzzy systems concern fundamental methodology to represent and process uncertainty and imprecision in the linguistic information. The fuzzy systems that use fuzzy rules to represent the domain knowledge of the problem are known as Fuzzy Rule Base Systems (FRBS). Literature suggests that effort in this respect appears to be quite rudimentary. In the present article, we propose a fuzzy rule guided novel technique that is functional devoid of any external intervention during execution. Experimental results suggest that this approach is an efficient one in comparison to different other techniques extensively addressed in literature. In order to justify the supremacy of performance of our proposed technique in respect of its competitors, we take recourse to effective metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE) and Peak Signal to Noise Ratio (PSNR).

Index Terms - Fuzzy Rule Base, Image Extraction, Fuzzy Inference System (FIS), Membership Functions, Threshold methods, Soft Computing, Fuzzy Image Processing, Feature based modeling

1 INTRODUCTION

In traditional computing methodology, the prime considerations are precision, certainty, and rigor. By contrast, the principal guidelines of soft computing [1] revolve around the following: tolerance for imprecision, uncertainty, partial truth and approximation. It will help to achieve tractability, robustness and low solution cost. This leads to the remarkable human ability of understanding distorted speech, deciphering sloppy handwriting, comprehending the nuances of natural language, summarizing text, recognizing and classifying images, driving a vehicle in dense traffic and, more generally, making rational decisions in an environment of uncertainty and imprecision. Soft computing is a consortium of methodologies that works synergetically and provides in one form or another flexible information processing capability for handling real life ambiguous situations. The guiding principle is to devise methods of computation that lead to an acceptable solution at low cost by seeking for an approximate solution to an imprecisely/precisely formulated problem. The theory of fuzzy logic [2] provides a mathematical strength to capture the uncertainties associated with human cognitive processes, such as thinking and reasoning.

The conventional approaches to knowledge representation lack the means for representing the meaning of fuzzy concepts. As a consequence, the approaches based on first order logic and classical probability theory do not provide an appropriate conceptual framework for dealing with the representation of commonsense knowledge, since such knowledge is by its nature both lexically imprecise and non-categorical. Fuzzy Logic is usually regarded as a formal way to describe how human beings perceive everyday concepts. In Fuzzy Image processing, fuzzy set theory [3] is applied to the task of image processing. Fuzzy Image Processing is depends upon membership values [4], rule-base and inference engine. Unlike classical logic systems, Fuzzy Logic (FL) aims at modeling the imprecise modes of reasoning, which is the human ability to make a rational decision when information is uncertain and imprecise. FL starts with the concept of a fuzzy set. A fuzzy set is a set without a crisp, clearly defined boundary. It can contain elements with only a partial degree of membership. Membership criteria are not precisely defined for most classes of objects normally encountered in the real world. A fuzzy set is characterized by a membership function, that takes values in the interval [0, 1], such that the nearer the value to unity, the higher the membership grade. The uncertainty in image extraction and subsequent segmentation from noise affected scene effectively handled by Fuzzy Logic. According to [5], fuzzy approaches for image segmentation can be categorized into four classes: segmentation via thresholding, segmentation via clustering, supervised segmentation and rule based segmentation. Among these categories,

- Koushik Mondal presently serves as Scientific Officer (System Administrator) in Indian Institute of Science Education and Research, Pune, India, E-mail: koushik@iiserpune.ac.in
- Paramartha Dutta currently engaged as Professor in Department of Computer and System Sciences in Visva Bharati University Santiniketan India. E-mail: paramartha.dutta@visva-bharati.ac.in
- Siddhartha Bhattacharyya presently serves as Associate Professor, RCC IIT Kolkata, E-mail: siddhartha.bhattacharyya@gmail.com

rule based segmentation are able to take advantage of application dependent heuristic knowledge and model them in the form of fuzzy rule base. In our case, the heuristic knowledge gathers by the process of already exist threshold segmentation methods that helped us to build the rule base. Thresholding is a simple shape extraction technique. If it can be assumed that the shape to be extracted is defined by its brightness, then thresholding an image at that brightness level should find the shape. Thresholding is clearly sensitive to change in illumination: if the image illumination changes then so will the perceived brightness of the target shape. Unless the threshold level can be arranged to adapt to the change in brightness level, any thresholding technique will fail. Its attraction is simplicity: thresholding does not require much computational effort. If the illumination level changes in a linear fashion, then using histogram equalisation will result in an image that does not vary. Unfortunately, the result of histogram equalisation is sensitive to noise: noise can affect the resulting image quite dramatically and this will help us to determine the minute changes in the original images clearly. Image Segmentation and subsequent extraction from noise-affected scene happen to be crucial phase of image processing. The complex process of human vision is yet not comprehensively explored, in spite of several decades of dedicated study on the problem, may it be from the perspective of basic science or from the viewpoint of research on intelligence. In computer vision, the complex process of recognizing shapes, colors, textures and subsequently grouping them automatically into separate regions or objects within a scene continues to be an open research avenue, intrinsically because of the uncertainty associated with it. Out of the twin objectives of segmentation and extraction, championed earlier, image segmentation appears to be a low-level image-processing task that aims at partitioning an image into regions in order that each region/ group consists of homogeneous pixels sharing similar attributes (intensity, colors etc.). The problem becomes even more challenging with the presence of noise in the image scene where the uncalled noise components need to be eliminated while preserving the image content as much as possible. Naturally, the extraction of objects prevalent in an image content from a noise affected background. The visual features such as shape, color and texture are extracted to characterize images in the phase of image extraction. Each of the features is represented using one or more feature descriptors. During the process, features and descriptors of the query are compared to those of the images in order to calculate rank of each indexed image according to its distance to the query. The extraction task transforms rich content of images into various content features. Feature extraction is the process of generating features to be used in the selection and classification tasks. Feature selection reduces the number of features provided to the classification task. Those features that are likely to assist in discrimination are selected and used in the classification task. Features that are not selected are discarded [6]. After the features are extracted, a suitable

classifier must be chosen. A number of classifiers are used and each classifier is found suitable to classify a particular kind of feature vectors depending upon their characteristics. The classifier used commonly is Nearest Neighbor classifier. The nearest neighbor classifier is used to compare the feature vector of the prototype with image feature vectors stored in the database. High-level feature extraction concerns finding shapes in computer images. To be able to recognise faces automatically, for example, one approach is to extract the component features. This requires extraction of, say, the eyes, the ears and the nose, which are the major face features. To find them, we can use their shape: the white part of the eyes is ellipsoidal; the mouth can appear as two lines, as do the eyebrows. Shape extraction implies finding their position, their orientation and their size. This feature extraction process can be viewed as similar to the way we perceive the world: many books for babies describe basic geometric shapes such as triangles, circles and squares. More complex pictures can be decomposed into a structure of simple shapes. Modular approaches partitions the classification task into some sub-classification tasks, solve each sub-classification task, and eventually integrates the results to obtain the final classification result. In other words, partitioning of the classification task is carried out such that each sub-problem can be solved in a module by exploiting the local uncertainties and exploiting the global uncertainties can combine the results of all the modules. The performance of each module can be improved by giving importance to the features based on their class discrimination capability for the output classes present in the module. In many applications, analysis can be guided by the way the shapes are arranged. The task of pattern classifier is to search the structure. This search becomes complicated because of the presence of uncertainties associated with the structure. Thus, the whole pattern classification process involves manipulation of the information supplied by the instances. The instances contain the information about the process generating them, and the extracted features reflect this information. The structures present inside the features represent the information in an organized manner so that the relationship among the variables in the classification process can be identified. Finally, in the last step, a search process recognizes the information from the structure. Now, if a new pattern is encountered, the machine detects the structure in which the input pattern belongs, and based on the structure the pattern is classified. Therefore, once the structure is found, the machine is capable of dealing with new situations to some extent. The issue of choosing the features to be extracted should be guided by the following concerns:

1. The features should carry enough information about the image and should not require any domain-specific knowledge for their extraction.
2. They should be easy to compute in order for the approach to be feasible for a large image collection.

3. They should relate well with the human perceptual characteristics since users will finally determine the suitability of the retrieved images.

In the other hand, two steps have to be considered in order to address any segmentation problem:

Step 1: Formalize the segmentation problem, a mathematical notion of homogeneity or similarity among image- regions need to be considered.

Step 2: An efficient algorithm for partitioning or clustering has to be derived particularly to carry the earlier step out in a computationally efficient manner.

The problems of image segmentation become more uncertain and severe when it comes to dealing with noisy images. The vagueness of image information arising out of admixture of the different components has been dealt with soft computing paradigm. Numerous articles and several surveys on gray /monochrome image segmentation techniques have to be reported in this regard [7] [8][11].

A formal definition of segmentation of an image can be defined as in [9]. Segmentation of image I is a partition P of I into a set of M regions $\{R_m, m=1, 2...M\}$ such that:

1. $\bigcup_{m=1}^M R_m = I$ with $R_m \cap R_n = \Phi$,
 $m \neq n, 1 \leq m, n \leq M$
2. $H(R_m) = true \forall m, 1 \leq m \leq M$
3. $H(R_m \cap R_n) = false \forall R_m$ and R_n
adjacent, $1 \leq m, n \leq M$

Figure 1: Segmentation definition

Here H is the predicate of homogeneity. A region is homogeneous if all its pixels satisfy the homogeneity predicate defined over one or more pixel attributes such as intensity, texture or color. On the other hand, a region is connected if a connected path exists between any two pixels within the region.

Because of the large diversity of segmentation methods, it is indeed difficult to exhaustively review each individual segmentation techniques up to now. However, segmentation methods can be broadly classified as [9][10][11][12][13]:

1. Region or boundary-based;
2. Graph-based;
3. Histogram-based;
4. Pixel based;

5. Area based;
6. Physics based;

This chapter is presented in the following manner. In the section 2, we would like to discuss survey of recent methodologies in this area; section 3 proposes our present work; section 4 clarifies the results and analysis followed by conclusion.

2 SURVEY

Gray scale image segmentation approaches are based on either discontinuity and/or homogeneity of gray level values in a region. The approach based on discontinuity, tends to partition an image by detecting isolated points, lines and edges according to abrupt changes in gray levels in two adjacent regions in the scene. The approaches based on homogeneity include thresholding, clustering, region growing and region splitting & merging. Several surveys are reported in the literature to this effect. Fu et al. discussed segmentation from the viewpoint of cytology image processing [7]. The paper categorized various existing segmentation techniques into three classes:

1. Characteristic feature thresholding or clustering
2. Edge detection and
3. Region extraction.

The segmentation techniques were summarized and comments were provided on the pros and cons of each approach. The threshold selection schemes based on gray level histogram and local properties as well as based on structural, textural and syntactic techniques were described [7][8][9]. Clustering techniques were regarded as "the multidimensional extension of the concept of thresholding". Some clustering schemes utilizing different kinds of features (multi-spectral information, mean/variation of gray level, texture, color) were discussed. Various edge detection techniques were presented, which were categorized into two classes - parallel and sequential techniques. The parallel edge detection technique [10][11] implies that the decision of whether a set of points is on an edge or not, depends on the gray level of the set and some set of its neighbors, which includes high emphasis on spatial frequency filtering, gradient operators, adaptive local operator, functional approximations, heuristic search and dynamic programming, relaxation and line & curve fitting, while the sequential techniques make decision based on the results of the previously examined points. A brief description of the major component of a sequential edge detection procedure was provided in [7][9]. In those papers region merging, region splitting and combination of region merging and splitting approaches briefly introduced. Haralick et al. classified image segmentation techniques into six major groups [8]:

1. Measurement space guided spatial clustering
2. Single linkage region growing schemes
3. Hybrid linkage region growing schemes

4. Centroid linkage region growing schemes
5. Spatial clustering schemes and
6. Split & merge schemes.

These techniques are compared on the problem of region merge error, blocky region boundary and memory usage. The hybrid linkage region growing schemes appear to be the best compromise between having smooth boundaries and few unwanted region merges. One of the drawbacks of feature space clustering is that the cluster analysis does not utilize any spatial information. The article also presented some spatial clustering approaches, which combine clustering in feature space with region growing or spatial linkage techniques. It provides a good summary of kinds of linkage region growing schemes. The problem of high correlation and spatial redundancy of multi-band image histograms and the difficulty of clustering using multi-dimensional histograms are also discussed. Sahoo et al. surveyed segmentation algorithms based on thresholding and attempted to evaluate the performance of some thresholding techniques using uniformity and shape measures [9]. It categorized global thresholding techniques into two classes:

- i. point-dependent techniques (gray level histogram based)
- ii. region-dependent techniques (modified histogram or co-occurrence based).

Histogram thresholding is one of the widely used techniques for monochrome image segmentation. It assumes that images are composed of regions distributed with different gray level ranges. As for color images, the situation is different from monochrome images because of presence of multiple features. Multiple histogram-based thresholding is able to decompose color space by thresholding histogram component-wise. Guo et al. adopted entropy based thresholding method [18]. Mode seeking is decided by the multi-modal probability density function (pdf) estimation and the mode can be found by thresholding the pdf. In the above approaches, thresholding is performed with only one color component at a time. Thus the regions extracted are not based on the information available from all three components simultaneously because the correlation among the three components is neglected. This problem can be solved if we can get hold of such an approach that the points in the 3D space are projected onto it and the projected points can be well separated. Generally, two or more characteristic features form a feature space and each class of regions is assumed to form a separate cluster in the space. The reason to use multiple characteristic features to perform image segmentation is that, sometimes, problems might crop up which are not solvable with one feature but is also solvable with multiple features. The characteristic features may be any features that could be used for the segmentation problem, such as the gray level value of multi-spectral images, gray level histogram, mean, deviation, texture, etc. Discussion on probabilistic relaxation and several methods of multi-thresholding techniques was also available in [8][11]. Spirkovska et al. regarded image segmentation in a machine vision system

as the bridge between a low-level vision subsystem including image processing operations (such as noise elimination, edge extraction etc.) to enhance the image quality on one hand and a high-level vision subsystem including object recognition and scene interpretation on the other [10].

Most gray level image segmentation techniques can be extended to color images, such as histogram thresholding, clustering, region growing, edge detection, fuzzy approaches and neural networks. Gray level segmentation methods can be directly applied to each component of a color space. The results can be combined in some way to obtain a final segmentation result. Segmentation may also be viewed as image classification problem based on color and spatial features [11]. Therefore, segmentation methods can be categorized as supervised or unsupervised learning /classification procedures. Power et. al. compared different color spaces (RGB, normalized RGB, HSI- hybrid color space) and supervised learning algorithms for segmenting fruit images [14]. Supervised algorithms include Maximum Likelihood, Decision Tree, K-Nearest Neighbor, Neural Networks, etc. Hance et al. explored six unsupervised image segmentation approaches [15]:

1. Adaptive thresholding
2. Fuzzy C-means (FCM)
3. SCT/center split
4. PCT (Principal Components Transform) median cut
5. Split and merge
6. Multi-resolution segmentation.

Some algorithms resort to combination of unsupervised and supervised methods to segment color images. Hu et al. used unsupervised learning and classification based on the FCM algorithm and nearest prototype rule [16]. The classified pixels are used to generate a set of prototypes, which are optimized using a multilayer neural network. The supervised learning is utilized because the optimized prototypes are subsequently used to classify other image pixels. Eom et al. employed a neural network for supervised segmentation and a fuzzy clustering algorithm for unsupervised segmentation [17]. Histogram thresholding is one of the widely used techniques for monochrome image segmentation. It assumes that images are composed of regions distributed with different gray level ranges. The histogram of an image can be separated into a number of peaks (modes) each corresponding to one region and there exists a threshold value corresponding to valley between the two adjacent peaks. However, there is limitation since all the existing thresholding techniques having notional resemblance to gray scale images.

In order to obtain the maximum information between two sources, mode (regions with high densities) and valley (regions with low densities), Guo et al. adopted entropy based thresholding method [18]. Mode seeking is decided by the multi-modal probability density function (pdf) estimation and the mode can be found by thresholding the pdf. A network for classifying an image

into distinct regions can be subjected to either supervised or unsupervised learning. The learning would be supervised if external criteria and/or intervention are used and matched by the network output otherwise the learning is unsupervised [19].

Genetic algorithm is another search strategy based on the mechanism of natural selection and group inheritance in the process of biological evolution [20][21]. It simulates the cases of reproduction, mating and mutation in sexual reproduction. GA looks each potential solution as an individual in a group (all possible solutions) and encodes each individual into an encoded domain where the genetic operators [21] can be effectively applied.

Fuzzy systems and Artificial Neural Network (ANN) are soft computing approaches to modeling expert behavior. The goal is to mimic the actions of an expert who solves complex problems. In other words, instead of investigating the problem in detail, one observes how an expert successfully tackles the problem and obtains knowledge by instruction and/or learning. Let us consider the significance of querying and rule generation, by referring to the example of decision making system generally followed up in medical diagnostics. The models are generally capable of dealing with non-availability of data, and can enquire the user for additional data when necessary. In the medical domain, for instance, data may be missing for various reasons; for example, some examinations can be risky for the patient or contraindications can exist, an urgent diagnostic decision may need to be made and some very informative but prolonged test results may have to be excluded from the feature set, or appropriate technical equipment may not be available. In such cases, the network can query the user for additional information only when it is particularly necessary to infer a decision. Again, one realizes that the final responsibility for any diagnostic decision always has to be accepted by the medical practitioner. So the physician may want to verify the justification behind the decision reached, based on personal expertise. This requires the system to be able to explain its mode of reasoning for any inferred decision or recommendation, preferably in rule form, to convince the user that the reasoning is correct. Human operators have an advantage over control and protection systems in terms of their experience and ability to assimilate a wide spectrum of information and new data. In contrast, computers have the advantage of being able to process such information much faster than their human counterparts. Fuzzy logic has that capability of taking advantage of the operators' experience and the fast data processing capability of computers. A learning process can be part of knowledge acquisition. In the absence of an expert or sufficient time or data, one can resort to reinforcement learning instead of supervised learning. In general, if one has knowledge expressed as linguistic rules, one can build a fuzzy system. On the other hand, if one has data or can learn from a simulation or the real task, ANN's are more appropriate [22]. Compared to

neural network model, fuzzy model integrate the knowledge representation and reasoning mechanism with the priori expert experience and knowledge, consistent with people's habits of mind, its structure and membership function parameters have obvious semantic meaning, it can be easily understood its internal operation mechanism by studying the rules of fuzzy system, in conclusion, the explanation is the most prominent feature of a fuzzy model. How to automatically construct the fuzzy systems with accuracy and proper explanation from data analysis, is the key point of this research area. The explanatory of Fuzzy classification system, so far has no clear definition, but is generally believed that the explanatory of fuzzy classification system is closely related with the number of the characteristics variables, the number of fuzzy rules, and the characteristics of membership functions, and the fuzzy classification system with fewer number of feature variables, fewer number of fuzzy rules has better explanation ability. In feature extraction [23], we generally seek invariance properties so that the extraction process does not vary according to chosen (or specified) conditions. That is, techniques should find shapes reliably and robustly whatever the value of any parameter that can control the appearance of a shape. As a basic invariant, we seek immunity to changes in the illumination level: we seek to find a shape whether it is light or dark. In principle, as long as there is contrast between a shape and its background, the shape can be said to exist, and can then be detected. It is clear that, any computer vision technique will fail in extreme lighting conditions; you cannot see anything when it is completely dark. Then, we often seek to find a shape irrespective of its rotation (assuming that the object or the camera has an unknown orientation): this is usually called rotation- or orientation invariance. Then, we might seek to determine the object at whatever size it appears, which might be due to physical change, or to how close the object has been placed to the camera. This requires size- or scale-invariance. These are the main invariance properties we shall seek from our shape extraction techniques. However, nature tends to roll balls under our feet: there is always noise in images. Also since we are concerned with shapes, note that there might be more than one in the image. If one is on top of the other, it will occlude, or hide, the other, so not all the shape of one object will be visible. But before we can develop image analysis techniques, we need techniques to extract the shapes. Extraction is more complex than detection, since extraction implies that we have a description of a shape, such as its position and size, whereas detection of a shape merely implies knowledge of its existence within an image [24]. In order to extract a shape from an image, it is necessary to identify it from the background elements. This can be done by considering the intensity information or by comparing the pixels against a given template. In the first approach, if the brightness of the shape is known, then the pixels that form the shape can be extracted by classifying the pixels according to a fixed intensity threshold. Alternatively, if the background image is

known, then this can be subtracted to obtain the pixels that define the shape of an object superimposed on the background. Template matching is a model-based approach in which the shape is extracted by searching for the best correlation between a known model and the pixels in an image.

There are alternative ways to compute the correlation between the template and the image. Correlation can be implemented by considering the image or frequency domains. Additionally, the template can be defined by considering intensity values or a binary shape.

Wang and Archer [26] have introduced ultrafuzzy sets for modeling decision-making under conflict, using a modified version of backpropagation. In case of ultrafuzzy sets, the membership function takes on fuzzy values. Ultrafuzzy interval of certainty factor is modeled as the consequent of a rule. Two fuzzy membership functions termed as participation and moderation functions, falling in the ultrafuzzy interval, are developed based on the well-known plausibility and belief functions [27]. The concept of plausibility and belief functions is used to construct conflict measures, which help in explaining the compromise phenomena observed in decision-making. This fuzzy decision-making model is capable of cumulating human knowledge and is claimed to be useful for maintaining consistency while making decisions. According to [28], It is appropriate to use FL when:

1. one or more of the variables are continuous and are not easily broken down into discrete segments;
2. a mathematical model of the process does not exist, or exists but is too difficult to encode;
3. a mathematical model of the process exists but is too complex to be evaluated fast enough for real-time operation;
4. high ambient noise levels are expected in the input signals; and/or
5. engineering interpretations become highly subjective and context dependent.

Engineering interpretations becomes highly subjective and context dependent when considering the additional information such as historical usage trends, weather, and system/component reliability data. Generally the rules and the membership functions used by the fuzzy logic for solving the classification problem are formed from the experience of the human experts. With an increasing number of variables, the possible number of rules for the system increases exponentially, which makes it difficult for experts to define a complete rule set for good system performance.

In fuzzy inferencing and rule generation approaches, the fuzzy classification rule described by Ishibuchi et al. [29], the partitioning is uniform, i.e., the regions continue to be split until a sufficiently high certainty of the rule, generated by each region, is achieved. Ishibuchi et al. extended this work later [30] by using an idea of sequential partitioning of the feature space into fuzzy

subspaces until a predetermined stopping criterion is satisfied and studied its application for solving various pattern classification problems. Wang and Mendel [31] developed a slightly different method for creating a fuzzy rule base, made up of a combination of rules generated from numerical examples and linguistic rules supplied by human experts. The input and output domain spaces are divided into a number of linguistic subspaces. Human intervention is sought to assign degrees to the rules and conflicts are resolved by selecting those rules yielding the maximum of a computed measure corresponding to each linguistic subspace. Rovatti and Guerrieri [32] have attempted to identify the correct rule structure of a fuzzy system when the target input-output behavior is sampled at random points. The assumption that a rule can either be included or excluded from the rule set is relaxed, and degrees of membership are exploited to achieve good approximation results. In [33], the proper fuzzy rule base for gray image extraction possessed with the help of correct rule structure. In the rule generation phase, different existing thresholding methods are used to build membership functions. Defuzzification methodologies are then used to extract well-behaving crisp rule sets. Symbolic minimization is carried out to obtain a compact structure that captures the high-level characteristics of the target behavior. For other details, one may refer to [35]–[36]. Dietterich et. al. [37] suggests a Multiple Classifier Systems that introduce a new way for building more accurate classifiers. He suggests three types of reasons namely, statistical, representational and computational, explaining why a classifier ensemble can be better than a single classifier. There are two approaches for making a classifier out of multiple classifier systems namely, classifier fusion and classifier selection [38]. In classifier fusion each classifier is supposed to know the whole data points in feature space, whereas in classifier selection each ensemble member is supposed to know one part of the feature space well and be responsible for objects in this part. In the selection approach we select one or more classifiers to label a new input. Some of the more famous methods for classifier fusion are Majority vote, Weighted Majority vote, Naive Bayes Combination, Multinomial Methods and Probabilistic Approximation [39]. The most commonly used method is assigning a competence to each classifier for the current input and choosing the most competent one. Kuncheva et. al. [38] shows that if we select the best classifier for each region in feature space, regardless of how we partition feature space, the resulting classifier is at least as good as the best classifier in the ensemble.

The theory of rough sets [39] has recently emerged as another major mathematical tool for managing uncertainty that arises from granularity in the domain of discourse, i.e., from the indiscernibility between objects in a set. The intention is to approximate a rough (imprecise) concept in the domain of discourse by a pair of exact concepts, called the lower and upper approximations. These exact concepts are determined by an indiscernibility relation on the domain, which, in turn,

may be induced by a given set of attributes ascribed to the objects of the domain. The lower approximation is the set of objects definitely belonging to the vague concept, whereas the upper approximation is the set of objects possibly belonging to the same. These approximations are used to define the notions of discernibility matrices, discernibility functions, reducts, and dependency factors, all of which play a fundamental role in the reduction of knowledge.

3 PRESENT WORK

It is interesting that all methods invariably performed poorly for at least one or two instances. Thus it was observed that any single algorithm could not be successful for all noisy image types, even in a single application domain. To obtain the robustness of the thresholding method, we explored the combination of more than one thresholding algorithm based on the conjecture that they could be complementary to each other. The combination of thresholding algorithms can be done at the feature level or at the decision level. At the feature level, we use, for example, some averaging operation on the maximum values obtained from individual algorithms; on the decision level, we have fusion of the foreground-background decisions, for example, by taking the majority decision. Thus it will help us on creating membership envelopes in the proposed system.

The algorithm for the proposed work is as follows:

Step 1. Read a noisy image as input

Step 2. Identify the Region of Interests of the image by different thresholding values

Step 3. Extract the image information in terms of pixel attributes and threshold values for future use.

Step 4. Construct the different membership envelopes of the input image.

Step 5. Generate fuzzy rules based on the numerical data obtained from the input image corrupted by noise. The fuzzy rule generation consists following steps:

- a. Discern Input and Output spaces into fuzzy regions
- b. Generate fuzzy rules from the given data
- c. Map the threshold values obtained from different methods in the corresponding fuzzy region
- d. Create a combined fuzzy rule base Determine a mapping on the basis of this combined fuzzy rule base.

Step 6. Approximate the value obtained in Step 5.

Step 7. Display the image constructed thus.

Fuzzy image processing is the collection of all approaches that understand, represent and process the images, their segments and features as fuzzy sets. Conventional approaches of pattern classification involve clustering training samples and associating clusters to given categories. Building classifiers involves capturing the similarity among the training patterns and assigning labels for the group of similar patterns. Capturing the similarity among patterns becomes complicated when a training pattern belongs to more than one class, i.e., the output classes are overlapping. Thus fuzzy uncertainty appears in form of similarity and overlap. Due to the lack of details, two input patterns may appear similar whereas the class labels may not be same. The complexity and limitations of previous mechanisms are largely due to the lacking of an effective way of defining the boundaries among clusters. This problem becomes more intractable when the number of features used for classification increases. On the contrary, fuzzy classification assumes the boundary between two neighboring classes as a continuous, overlapping area within which an object has partial membership in each class. This viewpoint not only reflects the reality of many applications in which categories have fuzzy boundaries, but also provides a simple representation of the potentially complex partition of the feature space. Both neural networks and fuzzy systems are dynamic, parallel processing systems that estimate input-output functions. They estimate a function without any mathematical model and learn from experience with sample data. A fuzzy system adaptively infers and modifies its fuzzy associations from representative numerical samples. In brief, we use fuzzy IF-THEN rules to describe a classifier. The representation and processing depend on the selected fuzzy technique and on the problem to be solved. A fuzzy system is comprised of five basic elements, as shown in Figure 2. A fuzzifier is responsible for mapping the crisp inputs from the system into fuzzy sets modeling the inputs. The second element is the knowledge base, which incorporates the required knowledge about the system in the form of fuzzy If-Then rules. The rules are governed by the relationships between the inputs and the way that they combine to produce the desired output. The third element is the fuzzy model, which is the group of fuzzy sets describing each of the input and output variables. The fuzzy sets partition the universe of discourse of a given input or output variable into a group of overlapping fuzzy sets. The fourth element is the fuzzy inference system, which is the reasoning process through which the fuzzified inputs are used to activate the relevant rules. The last element is the defuzzifier, which is the mechanism by which the fuzzy input set is converted into a single output value or control parameter.

The fuzzification and defuzzification steps are particularly important because of absence of any fuzzy hardware at our disposition. Therefore, the coding of image data (fuzzification) and decoding of the results (defuzzification) are steps indispensable that make possible to process images with fuzzy techniques. The

main power of fuzzy image processing lies in the effective use of the middle step (modification of membership values). Wang et al. [25] have used a fuzzy logic rule-based system to first determine a good feature set for the recognition of *Escherichia coli* O157:H7, a cause of serious health problems. Fuzzy membership functions are defined for each term set of each linguistic variable in the rules. The human inspired features of this reduced rule set are then incorporated in a multiple neural network fusion approach. The fuzzy integral is utilized in the fusion of the networks trained with different feature sets. Unfortunately, most of the available literature on rule generation does not provide such rigorous assessment on their pros and cons. There is also a preponderance of specific purpose techniques that are designed to work with a particular architecture. This limits the scope of comparing the various techniques in a single framework.

4 RESULTS AND ANALYSIS

The goal of this paper is to describe a generic system using a Mamdani rule base. Specifically, we are modeling the relationship among the images, its extracted counterpart and the fuzzy rule base system using as many as 15 well known thresholding methods. In pattern recognition and image processing, feature extraction is a special form of dimensionality reduction. Feature extraction is a general term for methods for constructing combinations of the variables, but still describes the data sufficiently accurately. All images of different categories can be distinguished via their homogeneousness or feature characteristics. All the thresholding methods are generally based on the characteristics of one or some features, which will help us to build an adaptive mechanism guided by some already established methods. We use Lena images having 256X256 and 512X512 dimensions with Gaussian and Salt and Pepper noise respectively. We are using different noise levels and use all the existing techniques and our proposed technique to extract images. The comparisons are listed in the TABLE-1 and TABLE-2. We evaluated all possible to measure of how well the rule described the actual system behavior over the domain where its antecedent was true. In this paper, we take proper care about how well a Mamdani rule base can be put to model the system, using rules that have high correctness.

For Gaussian noise, the corrupted image, subsequent result obtained by well-known methods and proposed fuzzy rule base method is depicted in Figure 3, Figure 4 and Figure 5 respectively and the same for the salt and pepper noise are depicted in Figure 6, Figure 7 and Figure 8 respectively. The comparison among different PSNR values of Gaussian and Salt and Pepper noises are illustrated in Figure 9 and Figure 10 respectively.

5 CONCLUSION

The main features and advantages of this approach are:

1. It provides us a general method to combine

measured numerical information into a common framework- a combined fuzzy rule base that theoretically entertains both numerical and linguistic information

2. It is a simple and straightforward single pass buildup procedure and hence is devoid of any time consuming iterative training as it happens in a comparable neural network or in a neuro-fuzzy approach
3. There is a lot of freedom in choosing the membership domains in the said design. In fact, this happens to be one of the fundamental challenges
4. This can be viewed as very general model free integrated fuzzy system for a wide range of image processing problems where "model free" means no mathematical model is required for the problem; "integrated" means the systems integrates all the reported threshold values that are integrated with the systems for finding ROIs and that can help to design adaptive fuzzy regions; and, "Fuzzy" denotes the fuzziness introduced into the system by linguistic fuzzy rules, fuzziness of data, etc.

There are two criteria used in assessing the quality of images. They are subjective criterion and objective criterion. The subjective criterion relies on human beings' individual judgment and interpretation. Naturally, it is shrouded with the possibility of inconsistency and lacks repeatability, it is also time consuming and expensive. One of the standard ways of subjective measurement is called Mean Opinion Score (MOS), it is very tedious, costly and could not be feasible in real time. It has five scales ranging from 'impairment is not noticeable' (best) to 'impairment is extremely objectionable' (worst). On the other hand, the objective criterion available relies on the result of computing some of the following statistical error based methods dependent on pixels difference. Overall image mean absolute error (MAE), overall image mean square error (MSE), signal-to-noise ratio (SNR), or peak signal-to-noise ratio (PSNR) figure this list. The smaller the MAE (or MSE) or the larger the SNR (or PSNR) is, the higher is the quality of the signal. It is fast and repeatable.

There is no universal theory on image segmentation yet that may be universally applicable in all types of images. This is because image segmentation is subjective in nature and suffers from uncertainty. All the existing image segmentation approaches are, in the main, ad hoc. They are strongly application specific. In other words, there are no general algorithms vis-à-vis color spaces that are uniformly good for all color images. An image segmentation problem is fundamentally one of psychophysical perception and it is essential to supplement any mathematical solutions by a priori knowledge about the image. The fuzzy set theory has

attracted more and more attention in the area of image processing because of its inherent capability of handling uncertainty. Fuzzy set theory provides us with a suitable tool, which can represent the uncertainties arising in image segmentation and can model the relevant cognitive activity of the human beings. Fuzzy operators, properties, mathematics, inference rules have found more and more applications in image segmentation. Despite the computational cost, fuzzy approaches perform comparable to or better than their crisp counterparts. The more important advantage of a fuzzy methodology lies in that the fuzzy membership function provides a natural means to model the uncertainty prevalent in an image scene. Subsequently, fuzzy segmentation results can be utilized in feature extraction and object recognition phases of image processing and subsequent computer vision. Fuzzy approach also provides a promising means for color image segmentation.

REFERENCES

- [1]. L. A. Zadeh, "Fuzzy logic, neural networks, and soft computing," *Communication ACM*, vol. 37, pp. 77-84, 1994.
- [2]. L.A. Zadeh, "Fuzzy Sets", *Information and Control*, 1965.
- [3]. L.A. Zadeh, "Outline of A New Approach to the Analysis of Complex Systems and Decision Processes", 1973.
- [4]. L.A. Zadeh, "Fuzzy algorithms," *Info. & Control*, Vol. 12, 1968, pp. 94-102.
- [5]. Bezdek J C, Kellar J, Raghu K, Pal N H, "Fuzzy models and algorithms for pattern recognition and image processing" Kluwar Academic Publishers, Boston 1999.
- [6]. IEEE Computer, Special issue on Content Based Image Retrieval, 28, 9, 1995.
- [7]. K. S. Fu and J.K. Mui, "A Survey on Image Segmentation", *Pattern Recognition*, 13, 3-16, 1981.
- [8]. R. M. Haralick, L. G. Shapiro, "Image Segmentation Techniques", *Computer Vision, Graphics and Image Processing*, 29, 100-132, 1985.
- [9]. P. K. Sahoo et. al., "A Survey of Thresholding Techniques", *Computer Vision, Graphics and Image Processing*, 41, 233-260, 1988.
- [10]. L. Spirkovska, "A Summary of Image Segmentation Techniques", NASA Technical Memorandum, 104022, June 1993.
- [11]. S.K. Pal et. al. "A Review on Image Segmentation Techniques", *Pattern Recognition*, 29, 1277-1294, 1993.
- [12]. E. M. Riseman and M. A. Arbib, "Computational Techniques in the Visual Segmentation of Static Scenes", *Computer Vision, Graphics and Image Processing*, 6, 221-276, 1977.
- [13]. W. Skarbek, A. Koschan, "Colour Image Segmentation - A Survey", Technical Report, Tech. Univ. of Berlin, Oct. 1994.
- [14]. W. Power and R. Clist, "Comparison of supervised learning techniques applied to color segmentation of fruit image", *Proceedings of SPIE, Intelligent Robots and Computer Vision XV: Algorithms techniques, Active Vision and Materials Handling*, 370-381, Nov. 19-21, 1996, Boston, Massachusetts.
- [15]. G.A. Hance, S. E. Umbaugh, R. H. Moss and W. V. Stoecker, "Unsupervised color image segmentation with application to skin tumor borders", *IEEE Engineering in medicine and biology*, 104-111, January/February, 1996.
- [16]. J. Hu, H. Yan, A. N. Chalmers, "Color image segmentation using fuzzy clustering and supervised learning", *Journal of Electronic image*, 3(4), 397-403, Oct. 1994.
- [17]. K. B. Eom, "Segmentation of monochrome and color texture using moving average modeling approach", *Image Vision Computing*, 17(1999), 233-244.
- [18]. G. D. Guo, S. Yu and S. D. Ma, "Unsupervised Segmentation of Color Images", *IEEE International Conference on Image Processing*, 299-302, 1998.
- [19]. S. Bhattacharyya, P. Dutta, U. Maulik, "Binary object extraction using bi-directional self-organizing neural network (BDSONN) architecture with fuzzy context sensitive thresholding", *International Journal on Pattern Analysis and Applications*, Springer-Verlag, 2010.
- [20]. D. E Goldberg, 1989, "Genetic Algorithm in Search, Optimization and Machine Learning", Addison-Wesley Reading, Massachusetts.
- [21]. B. Bhanu, S. Lee, "Genetic Learning for Adaptive Image Segmentation", Kluwar Academic Press, 1994.
- [22]. D. Nauck, F. Klawonn, and R. Kruse, *Foundations of Neuro-Fuzzy Systems*. Chichester, U.K.: Wiley, 1997.
- [23]. Mark S. Nixon, Alberto S. Aguado, "Feature extraction and Image Processing" Newnes, Reed Elsevier plc. Group, 2002.
- [24]. R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image processing using MATLAB*. Upper Saddle River, N. J.: Pearson Prentice Hall, 2004.
- [25]. D. Wang, J. M. Keller, C. A. Carson, K. K. McAdoo-Edwards, and C.W. Bailey, "Use of fuzzy-logic-inspired features to improve bacterial recognition through classifier fusion," *IEEE Trans. Syst., Man, Cybernetics*, vol. 28, pp. 583-591, 1998.
- [26]. S. Wang and N. P. Archer, "A neural network based fuzzy set model for organizational decision making," *IEEE Trans. Syst., Man, Cybernetics*, vol.28, pp. 194-203, May 1998.
- [27]. G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.
- [28]. E. Cox, "Fuzzy fundamentals," *IEEE Spectrum*, vol. 29, no. 10, pp. 58-61, Oct. 1992.
- [29]. H. Ishibuchi, K. Nozaki, and H. Tanaka, "Distributed representation of fuzzy rules and its application to pattern classification," *Fuzzy Sets System*, vol. 52, pp. 21-32, 1992.
- [30]. H. Ishibuchi, K. Nozaki, and H. Tanaka, "Efficient fuzzy partition of pattern space for classification problems," *Fuzzy Sets Syst.*, vol. 59, pp. 295-304, 1993.
- [31]. L. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. Syst., Man, Cybernetics*, vol. 22, pp. 1414-1427, 1992.
- [32]. R. Rovatti and R. Guerrieri, "Fuzzy sets of rules for system identification," *IEEE Trans. Fuzzy Syst.*, vol. 4, pp. 89-102, 1996.
- [33]. Koushik Mondal, Paramartha Dutta, Siddhartha Bhattacharyya, "Gray image extraction using fuzzy Logic" *IEEE Second International Conference on Advanced Computing & Communication Technologies*, pp.289-296,2012.
- [34]. S. Abe and M. S. Lan, "Fuzzy rules extraction directly from numerical data for function approximation," *IEEE Trans. Syst., Man, Cybernetics*, vol. 25, pp. 119-129, 1995.
- [35]. T. Hong and C. Lee, "Induction of fuzzy rules and membership functions from training examples," *Fuzzy Sets Syst.*, vol. 84, pp. 33-37, 1996.
- [36]. "Special issue on fuzzy diagnosis," *Artificial Intelligence Med.*, vol. 16, no. 2, 1999.
- [37]. T. G. Dietterich, "Ensemble methods in machine learning," In J. Kittler and F. Roli, editors, *Multiple Classifier Systems*, volume 1857 of *Lecture Notes in Computer Science*, Cagliari, Italy, 2000, Springer, pp. 1-15.
- [38]. Ludmila I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, John Wiley & Sons, 2004, ISBN 0-471-21078-1.
- [39]. Z. Pawlak, *Rough Sets, Theoretical Aspects of Reasoning about Data*. Dordrecht, The Netherlands: Kluwer, 1991.

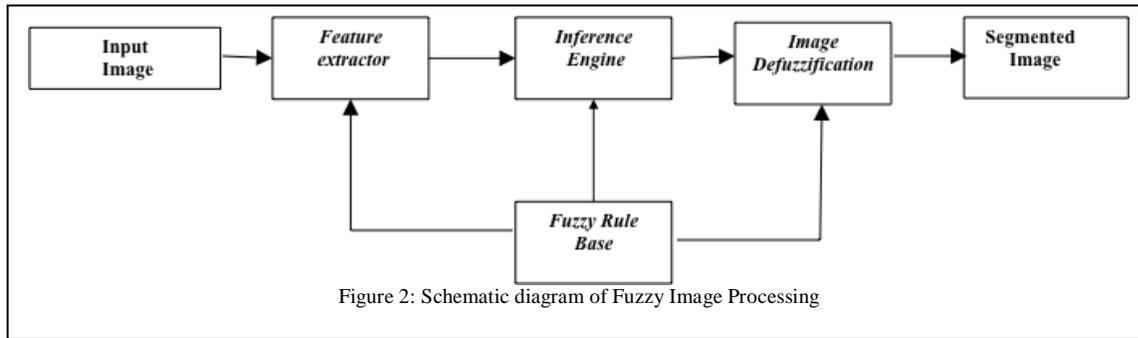


TABLE 1: COMPARISON OF GAUSSIAN NOISE REDUCTION WITH THE HELP OF DIFFERENT THRESHOLDING METHOD AND OUR PROPOSED METHOD

PSNR Calculation for Gaussian Noise					
Threshold method/Sigma	15	30	45	60	75
Default	18.7977	18.3826	17.8666	17.3449	16.9336
Huang	18.6423	18.3849	17.8578	17.3449	16.9263
Iso Data	18.9834	18.6435	17.4262	17.2113	16.9321
Li	18.6785	18.3245	17.8166	17.2354	16.8386
Max Entropy	18.7496	18.6748	17.4992	17.3166	16.9392
Mean	18.6342	18.3242	17.6822	17.3166	17.1336
Min Error	18.9321	18.3449	17.9321	17.8866	16.9336
Minimum	18.2435	18.1262	17.4221	17.3971	16.9491
Moments	18.9213	18.6314	17.8578	17.6314	16.937
Otsu	18.9932	18.6808	17.5817	17.4262	16.9336
Percentile	18.9213	18.6314	17.4213	17.3376	16.9321
RenyiEntropy	18.9491	18.6718	17.6166	17.4262	16.9392
Shanbhag	18.7661	18.2381	17.9932	17.7977	17.6166
Triangle	8.998	8.6314	7.7262	7.3216	5.4422
Yen	18.4262	18.2166	17.7143	17.4132	16.9402
Proposed method	29.6435	29.3126	28.9962	28.7143	28.6166



Figure 3: Image corrupted by Gaussian Noise



Figure 4: Image extracted by proposed method

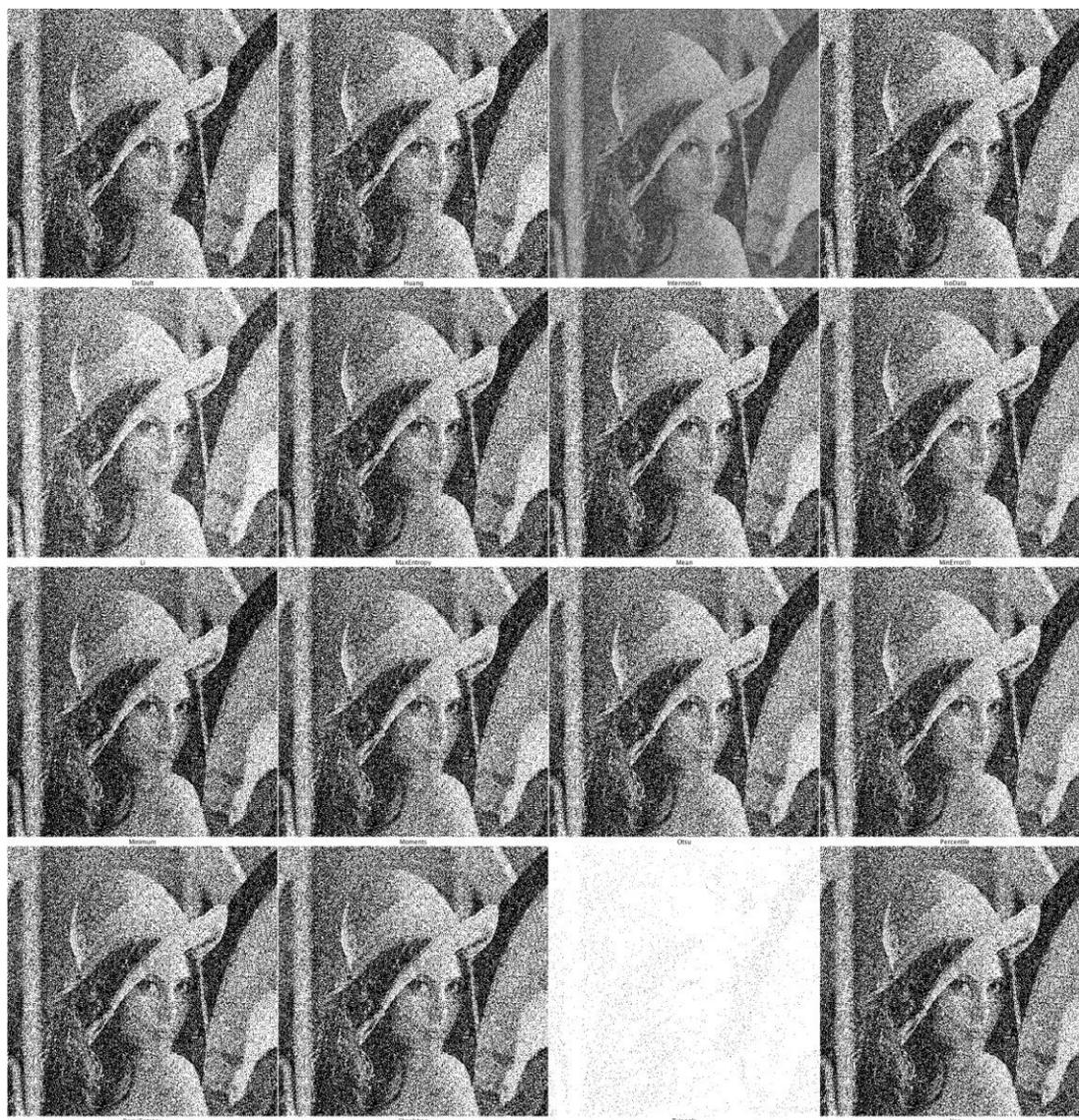


Figure 5: Image extraction by different thresholding methods

TABLE 2: COMPARISON OF SALT AND PEPPER NOISE REDUCTION WITH THE HELP OF DIFFERENT THRESHOLDING METHOD AND OUR PROPOSED METHOD

PSNR Calculation for Salt and Pepper Noise						
Threshold method/%	10	20	30	40	50	60
Default	18.7933	18.4693	18.1202	17.7935	17.4915	17.2249
Huang	17.1179	16.5558	16.0381	16.2808	15.9843	15.8381
Iso Data	18.5741	18.4853	18.1272	17.792	17.4915	17.2249
Li	17.7779	17.4125	17.0988	16.7947	16.6229	16.4197
Max Entropy	19.0767	18.4389	17.9857	17.6086	17.324	17.0689
Mean	18.5855	18.2576	17.927	17.6422	17.3936	17.1514
Min Error	15.4763	14.4525	14.452	14.5692	14.5671	14.6786
Minimum	15.7978	15.802	15.8024	15.8438	16.6312	16.5642
Moments	19.009	18.5407	18.1272	17.7814	17.4832	17.2228
Otsu	18.8227	18.4853	18.1377	17.7922	17.4869	17.2168
Percentile	18.4624	18.0417	17.6974	17.4219	17.1763	16.952
RenyiEntropy	18.9433	18.3988	17.946	17.6086	17.324	17.0847
Shanbhag	18.6629	18.23	17.8106	17.5152	17.2395	17.0081
Triangle	8.385	8.0863	5.811	5.2963	5.4104	5.5636
Yen	18.7309	18.1383	17.7587	17.4427	17.1979	16.9721
Proposed method	23.7809	23.6086	23.2963	22.9447	22.5481	22.4017



Figure 6: Image corrupted by Salt and Pepper Noise



Figure 7: Extracted image by Proposed Technique

The extracted images of 40% Salt and Pepper noise using different techniques are as follows:



Figure 8: Image extraction by different thresholding methods

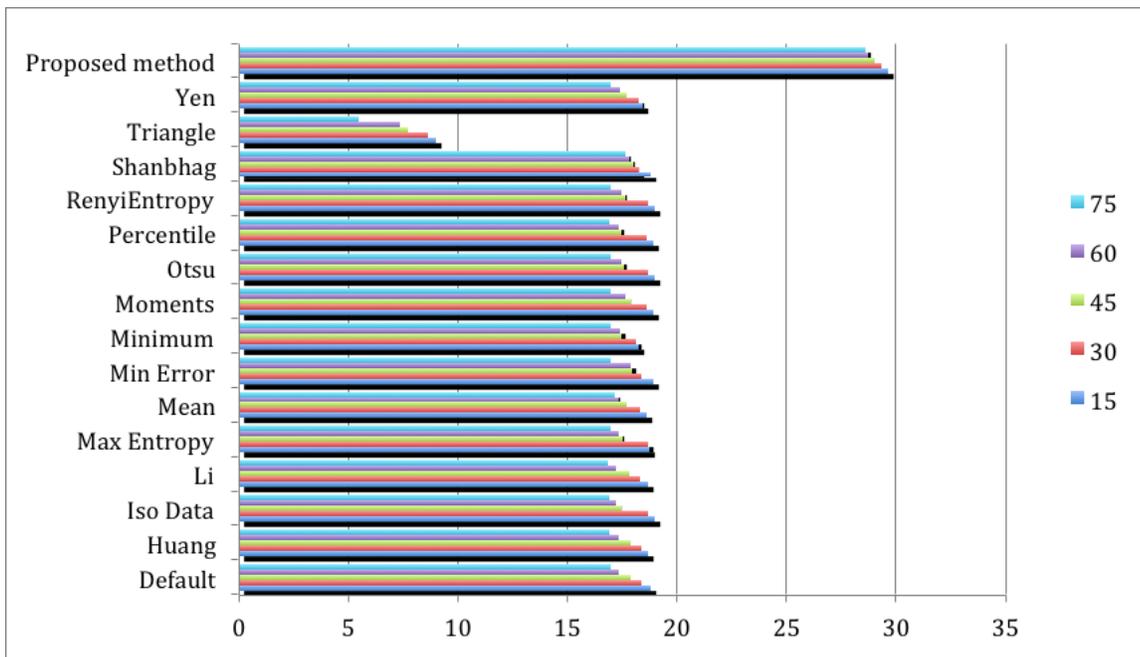


Figure 9: PSNR Calculation for Gaussian Noise

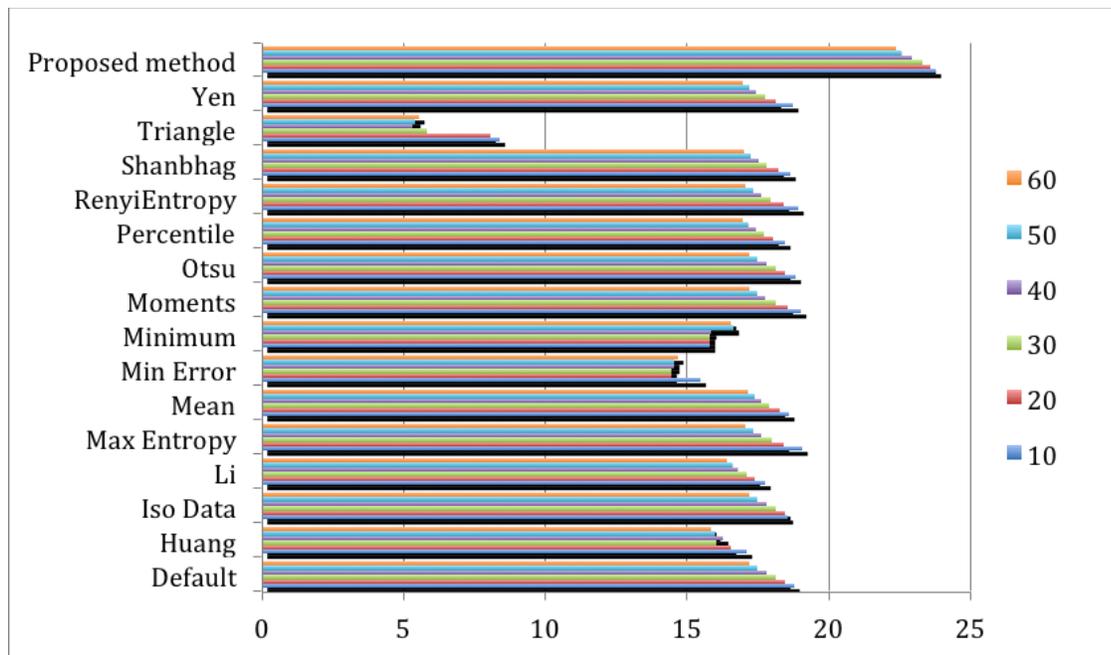


Figure 10: PSNR Calculation for Salt and Pepper noise

Effects of Tool Rake Angle on Tool Life in Turning Tools

Engr. Kaisan Muhammad Usman

ABSTRACT- In this work, the effect of tool rake angles on tool life was studied, the rake angles of 0° , 5° , 10° , 15° , and 20° and a constant clearance (Relief angle) of 8° were used to turn bright mild steel on the lathe machine, with a high speed steel of 18mm side as cutting tool and soluble oil was used as coolant. This is all in order to explore the energy savings opportunities during regrinding of tools, useful production time and energy is being wasted due to regrinding or re-sharpening of tools when cutting tools got worn or blunt, selection of the best rake angle which elongates tool life goes a long way in saving these time and energy. It was observed that, the rake angle of 20° gave the longest tool life as well as the best surface finish and yielded continuous chips formation.

Index Terms— Bright Mild Steel, Rake Angles, Relieve Angle, Wedge Angle, Tool Life, Turning Tools, Surface Finish

1 INTRODUCTION

Useful production time and energy are being wasted during regrinding or re-sharpening of cutting/turning tools during machining operations. Machining is one of the methods used in manufacturing engineering. Other methods include: casting, forming, grinding, shaping, finishing etc. The quest for profit maximization in manufacturing process makes it necessary for engineers and scientist to explore the optimum processes which requires less time and minimum energy for maximum output. Cutting tools are the basic tools used in machining operations which involve turning, milling, grinding, drilling, boring, planning and shaping and therefore tool life has to be elongated as much as possible in order to save time, energy and resources for optimum profit generation by a production firm. On this account, it has become necessary for us to investigate the effects of cutting parameters on tool life. According to Parson's, (1966) in 1907, Taylor has indicated cutting speed as the major determinant of tool life.

1.1 DEFINATIONS OF TERMS

Rake Angle (α): This is the angle between the tool face and the plane normal to the surface of the cut and pressing through the tool cutting edge (Edwards, 1993).

Clearance angle (β): This is also known as relief angle, it is the angle between the tool face adjacent to the

surface of the cut. This angle helps to eliminate rubbing between the tool and the surface being cut and hence reduces friction to the beeriest minimum. According to Charles' (1971) a clearance angle of 6° to 8° is large enough to prevent excessive rubbing of the tool on the work piece.

Wedge Angle (γ): This is the angle formed by the rake and the relief angle, which is between the rake face and the relief face or between the tool face and the tool relief face.

In general:

$$\alpha + \beta + \gamma = 90^\circ \quad (1)$$

Other useful angles: Other relevant angles are: Share angle, end relief angle, approach angle, end rake angles, nose angles, end cutting angle and the angle of inclination to the main cutting edge.

Tool Life: This is the period of cutting with the tool measured between regrinding. It is simply the time between regrinding or re-sharpening of tools. It refers to the continuous cutting period, regrinding is not generally delayed until the tool has broken down completely, the tool might have to be ground for example, because the quality of the surface finish is deteriorating or there is tendency of the tool not to maintain dimensional accuracy and so on.

• *Engr. Kaisan Muhammad Usman is a Scientific Officer with Department of Energy Management, Training and Manpower Development, Energy Commission of Nigeria and currently pursuing masters degree program in Energy System Engineering in Ahmadu Bello University, Zaria, Nigeria, PH-+2348036943404. E-mail: engineer.kaisan@gmail.com*

1.2 MODES OF TOOL FAILURES

There are several ways in which tools fail; tool life is determined as soon as any of the failures is encountered. To elongate tool life, the rate of such failures must be brought to minimum. The following are the modes of failure:

Tool Breakage

Excessive Tool Wear

Rough Surface Finish of Work

Tool Chatter

Over heating

Tool wear is the gradual or the progressive wearing of certain regions of the face or flank of the cutting tool. The three types of wear commonly identified during metal cutting are:

Adhesion

Abrasion and

Diffusion

1.3 COOLANTS

In many cutting operations, fluids are used to cool and lubricate both the cutting tool and the work piece. Such fluids are called coolants. Cooling increases tool life and helps to stabilize the side of the finished parts. Lubrication reduces friction thus, decreases the heat generated and the power required for a given cut. Cutting fluids include water based solutions, chemically inactive oils and synthetic fluids.

2.0 LITERATURE REVIEW

Mustafa Günay et al (2004) investigated the effect of cutting tool rake angle on main cutting force, for machining rotational parts by sharp cutting tools. A special dynamometer was designed and produced to measure the forces for this purpose.

Similarly, Erry Yulian T. Adesta et al (2009) investigated tool wear and surface finish in high speed

turning using cermet insert by applying negative rake angles, The experimental results showed that by increasing negative rake angles the higher wear occurred shorter duration of tool life and poor surface finish. The different between high Tool Wear and Surface Finish Investigation in High Speed Turning Using Cermet Insert by Applying Negative Rake Angles 181 speed and conventional speed was also observed in that experiment. High speed turning gives shorter tool life, high wear rate but finer surface finish than conventional one.

Alexius Anak An'yan, (2008) worked on to study the effect of grinding process parameters namely depth of cut, number of passes, and use of coolant on grinding force of aluminium alloy (AA6061-T6). A three component force transducer dynamometer (Kistler Model Type 5070) was used to measure grinding forces in this experiment. A full factorial experimental design was used as the approach for the design of experiment. Through the analysis of variance (ANOVA) conducted, it was found that the most significant parameter is the usage of coolant followed by depth of cut.

Viktor P. Astakhov and J. Paulo Davim (2008), worked on Tools (Geometry and Material) and Tool Wear, presents the basic definitions and visualisations of the most important components of the cutting tool geometry essential in the consideration of the machining process. The types and properties of modern tool materials are considered as well, as a closely related topic, as these properties describe to a great extent the restrictions on tool geometry. The basic mechanisms of tool wear are discussed. Criteria and measures of tool life are also considered in terms of Taylor's tool life models as well as in terms of modern tool life assessments for cutting tools used on computer numerical control (CNC) machines, manufacturing cells and production lines.

Finally, Hamidreza Shahabi Haghighi, (2008) in PhD theses, studied the direct effect of tool nose wear which is in contact to the surface profile of work piece directly, on the surface roughness and dimensional deviation of work piece using a developed machine vision in finish turning operation.

In this work however, the effects of tool rake angle on tool life in turning tools was investigated, the details of

the experimental procedure is elaborated in the next chapter.

3.0 MATERIALS AND METHODS

3.1 LIST OF MATERIALS/ EQUIPMENT

The following Machines, materials and apparatus were used in the experiment to investigate the effects of tool rake angles on tool life during turning operations:

Centre Lathe Machine

Universal Grinding Machine

Power Saw Machine

High Speed Steel Turning Tools 18mm, 8° clearance angle and 12° end relief angle

Bright Mild Steel (SAE 1025, Composition: 0.25%C, 0.4%Mn, 0.04%P, 0.05%S)

Protractors

Lanterns

Coolant (Soluble Oil)

Vanier Scale

Grease for Lubrication of Centre Drill

Stopped watch

Record Sheet and Pen

3.2 PREPERATION OF CUTTING TOOLS

The high speed steel cutting tools were ground on the universal grinding machine; various rake angles of 0°, 5°, 10°, 15°, and 20° were obtained. The clearance angle of 8° was made for all the corresponding rake angles on each HSS turning tool. After complete grinding, the protractor was used to measure the accuracy of all the angles obtained and all were found satisfactory. For 0°, no need of any grinding, rather the most flat and smooth HSS side was adopted for the purpose.

3.3 WORKPIECE PREPARATIONS

Using the power saw, the bright mild steel bar of 80mm diameter was cut into shorter lengths 200mm lengths to allow for easy mounting on the centre lathe machine, the coolant was used during the cutting process to minimise heating of both the power saw blade and the work pieces. The cylindrically shaped work pieces were obtained.

3.4 TOOL LIFE DETERMINATION

Using 3-jaw chucks, one of the bright mild steel work pieces was mounted on the head stock of centre lathe machine. The work piece was then centre drilled using a small drilled spindle which was attached to the tail stock. The drilled hole was then re-bored using a drilled of larger diameter, and grease was applied to the hole.

The HSS cutting tool of 18mm side, 0° rake angles and 8° clearance angles was used to turn the work piece. The whole cutting process was cooled continuously with soluble oil.

The cutting process carried out with the following selected conditions:

Depth of cut = 2mm

Cutting Speed = 250 rev/min

The lathe machine was then powered up and the tool was engaged on the work piece, the stopped watch was started as soon as the tool engaged the work piece, as cutting has almost started, the cutting was continued until the tool seized to cut or failed completely, the time taken was measured and recorded. Other distinguished parameters were also recorded during the experiment; these were the length of bar turned, initial and final diameters off the work piece and types of the chips produced.

The whole experiment was repeated using the other tools of 5°, 10°, 15°, and 20°, with clearance angles and all other conditions being maintained as in the results are presented in Table 4.2 below.

4.0 RESULTS AND DISCUSSIONS

The result of the experiment described in the 3.4 above is presented in the table 4.2, from the table, it can be seen that there are variations in the tool life as well as

length of the bar turned when the rake angle varied. The variation in the length of the bar turned makes it possible for us to determine the rates of metal removal for each rake angle. The volume of metal removed is given by:

$$V_m = Ctl \tag{2}$$

Where:

- V_m = volume of metal removed
- C = circumference of the work piece
- t = depth of cut

l = length of the bar turned

$$\text{But, } C = \pi D \tag{3}$$

Where:

D = diameter of the work piece

The volume of the metal removed is one of the most important variables in determining the best rake angle during machining operations. Under ideal cutting conditions the best rake angle is expected to yield the highest volume of metal removed before tool failure.

TABLE 4.2

EFFECTS OF TOOL FACE ANGLES ON TOOL LIFE AND VOLUME OF METAL REMOVED

SAMPL ES	RAKE ANGLE(α^0)	CLEARANCE ANGLE (β^0)	WEDGE ANGLE (γ^0)	LENGHT OF CUT (mm)	TOOL LIFE (min)	VOLUME OF METAL REMOVED (mm^3)
1	0	8	82	127.0	6.4	63845.550
2	5	8	77	85.8	3.4	43133.376
3	10	8	72	143.5	6.6	72140.320
4	15	8	67	63.0	4.4	31671.360
5	20	8	64	283.0	13.5	142269.760

The results presented in the table 4.2 above are presented in the figures below:

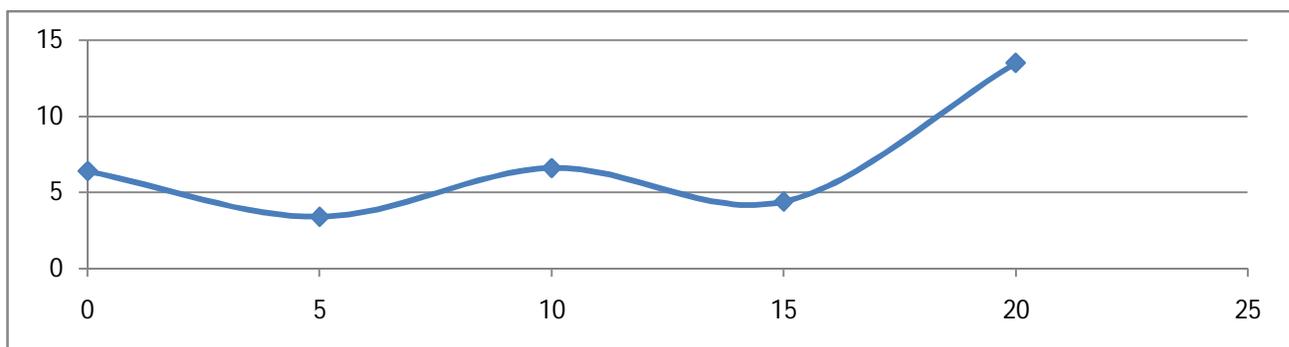


Figure1 Effects of Tool Rake Angles on Tool Life

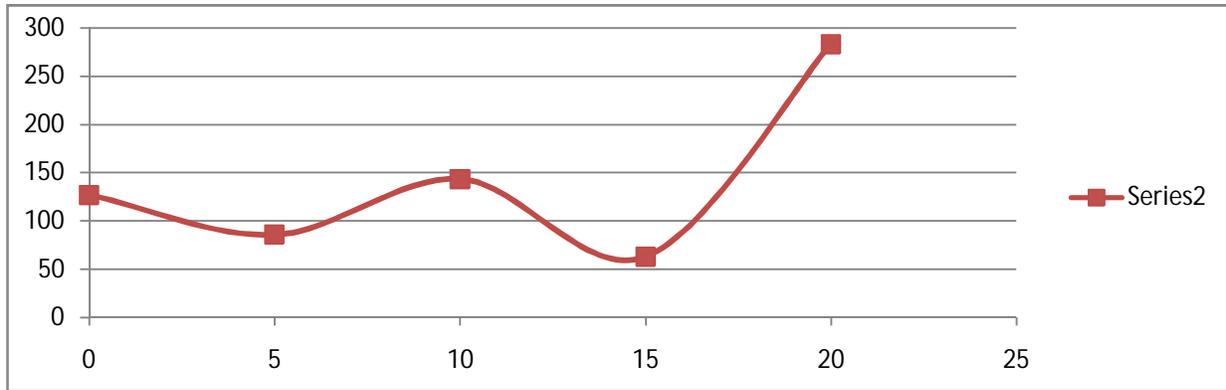


Figure2 Effects of Rake Angle on Length of Cut

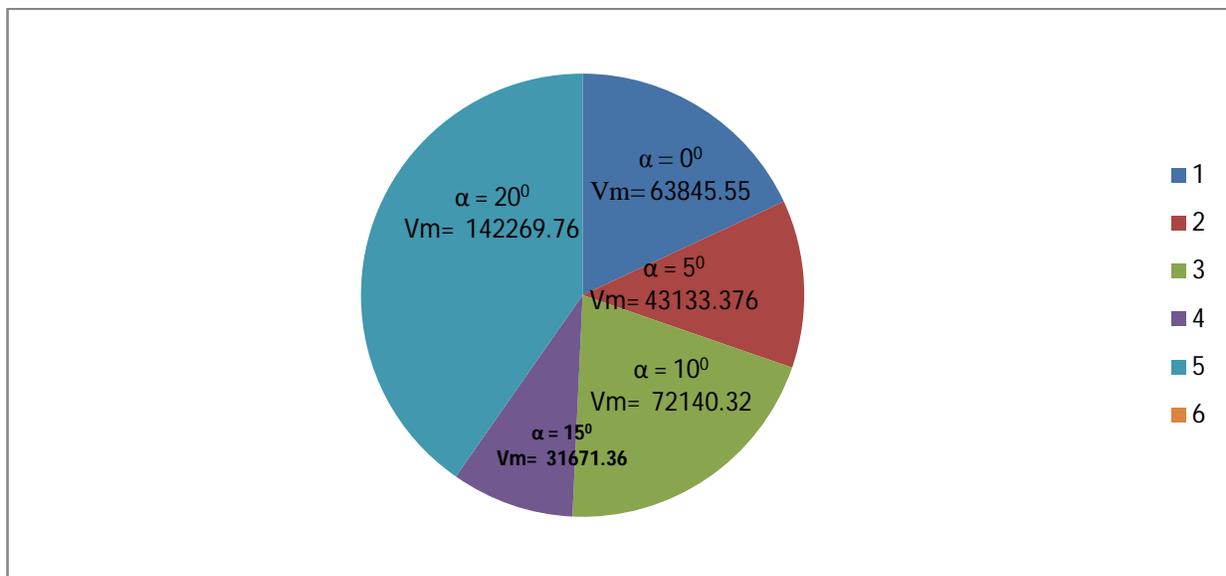


Figure 3 Effects of Rake Angle on Volume of Metal Removed

From the figures above, it would be clearly viewed that the rake angle of 20° gave the longest tool life of 13.5mins and the highest volume of metal removed of 142269.76 mm³ and a close observation on the work pieces shown that, only the work piece worked upon by the 20° rake angle yielded a continuous chips formation, which is also an evidence of good cutting conditions. The time and energy wasted during regrinding or re-sharpening of tools will be saved using the best rake angle. Time and energy are very important parameters used in optimizing the production capacity of any manufacturing firm, as such, we have to be careful with wasting such

important parameters unnecessarily more importantly during machining operations, whence, selection of the best rake angle in turning is very essential.

5.0 CONCLUSION

The tool face angles have effects on tool life in turning tools more specially during turning operations. While machining (Turning) bright mild steel with HSS turning tools, a rake angle of 20 gave the longest tool life, yielded continuous chips formation and produced the best surface finish.

6.0 REFERENCES

1. Charles R. H. (1971), Machine Tools and Process for

- Engineers, McGraw Hills Inc., Chapter Eight and Ten
2. Edwards R. (1993), Cutting Tools, The University Press, Cambridge, Chapter Five
 3. Persons B. A. J. (1966), Production Tooling Equipment, Macmillan and Company Limited, Chapters Eleven and Twelve
 4. Hamidreza Shahabi Haghighi (August, 2008), Study on The Effect of Tool Nose Wear on Surface Roughness and Dimensional Deviation of Work piece in Finish Turning Using Machine Vision, Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy, Malaysia
 - 5.
 6. Viktor P. Astakhov and J. Paulo Davim, Tools (Geometry And Material) And Tool Wear, General Motors Business Unit of PSMi, 1255 Beach Ct., Saline MI 48176, USA., Department of Mechanical Engineering, University of Aveiro, Campus Santiago, 3810-193 Aveiro, Portugal, 2008, <http://www.springer.com/978-1-84800-212-8>
 - 7.
 8. Alexius Anak An'yan, Effect of Grinding Process Parameters on Grinding Force of
 9. Aluminium Alloys (Aa6061-T6), A final year report submitted in partial fulfillment of the requirements for the
 - award of the degree of Bachelor of Mechanical Engineering with Manufacturing Faculty of Mechanical Engineering, Universiti Malaysia Pahang, November 2008
 - 10.
 11. Mustafa G'unay, Ihsan Korkut, Ersan Aslan, Ulvi S,eker, Experimental investigation of the effect of cutting tool rake angle on main cutting force, Journal of Materials Processing Technology 166 (2005) 44-49 Received 10 September 2003; received in revised form 22 December 2003; accepted 20 July 2004
 - 12.
 13. Erry Yulian T. Adesta, Muhammad Riza, Muataz Hazza, Delvis Agusman and Rosehan, Tool Wear and Surface Finish Investigation in High Speed Turning Using Cermet Insert by Applying Negative Rake Angles, European Journal of Scientific Research, ISSN 1450-216X Vol.38 No.2 (2009), pp.180-188 © EuroJournals Publishing, Inc. 2009, <http://www.eurojournals.com/ejsr.htm>
 - 14.
 15. K. V. Santha Kumari, Dipak Ranjan Jana and Anjani Kumar, Effects of Tool Setting on Tool Cutting Angle on Turning Operation, ARPN Journal of Engineering and Applied Sciences, VOL. 5, NO. 5, MAY 2010, ISSN 1819-6608, www.arpnjournals.com

Detecting Pornography on Web to Prevent Child Abuse – A Computer Vision Approach

Nilesh J.Uke, Dr. Ravindra C. Thool

Abstract—This paper explores the sensitive area from social, ethical and technical perspective and discuss an automated computer vision system for identifying whether there is a human nude present in a video which is browsed on the internet. Pornography is neglected problem in the society and in researchers due to its sensitive issue. But it cannot be ignored due to Internet reaching every child in the society. Current methods are not sufficient and enough to track child pornography. Proposed system consists of three phases. Segmentation phase accept the input video stream and extract the key frames which are candidates for nude images detection. Amount of nudity will be detected from rapid moving object detection phase. In classification phases segregation of objectionable video will be marked porn or non-porn depending upon the judgment criteria, decided on multiple the audio and video features.

Index Terms— Computer Vision, Object Detection, OpenCV, Segmentation, Skin Pixel, Video Classification

1 INTRODUCTION

THE word “pornography” comes from the Greek “pornographos” literally meaning writing about prostitutes.

Apart from this one of the commonly accepted definitions of “pornography” in modern times defines it as sexually explicit material (verbal or pictorial) that is primarily designed to produce sexual arousal in viewers [1].

In very short period of time, internet has become readily available to educational institutes, organizations and homes. According to survey conducted by Internet and Mobile Association of India (IAMAI) and IMRB India’s Internet users crossed 100 million in Sep 2011 a growth of 13% against last year. In India children account for a forty two percent of the country’s population, have long been the victims of some of the most brutal sexual crimes. This issue is always on top most priority for law enforcement agencies since the beginning of digital trafficking of child pornography.

Images and videos are essential part of today’s Internet. Basically they are intended to capture user’s attention and make effective interfaces. But now they are contributing towards harmful contents to the child and teens. This brings a new risk to younger generation and to the parents. Indian Government along with law enforcement agencies are trying to develop strong legal framework to deal with cases of child sexual abuse. Filtering sensitive contents has become a field of research of computer vision, artificial intelligence and multimedia field. Currently for detecting offensive material and the individuals responsible remain relatively primitive, requiring much human intervention and time. There is an urgent need for parents and children of today to have tools like filtering software to prevent people, especially children, from adult

images watching effectively and efficiently. One of the most difficult challenges is the problem of identifying pornographic images and videos is its semantic content. Simple solution to this problem can be performed using IP address blocking of an obscenity site.

2 SIGNIFICANCE OF RESEARCH

The main objective of this research is to formulate a new stout method of video interrogation to classify as pornographic. Our trust is to have a low false positive and low false negative detection rate in multiple frames in the video, accuracy of detection of video being the primary objective for law enforcement agencies. Essentially pornographic video scene detection system is a kind of classification problem.

2 REVIEW OF EXISTING METHODOLOGY

Many researchers have contributed by developing different algorithm to detect adult images and videos. Most of the work related to pornography detection is based on images. Brief survey on web image content can be found in [2]. Forsyth et al. proposed a body plan that detected different parts of the body by their geometric constraints for detecting the naked people. Their system marks skin - like pixels using combined color and texture properties. These skin regions are then fed to a grouper used for grouping human figure using geometric constrains [3].

Lv and Yang applied change in optical flow volume to extract important key frames and later used skin color detection method to detect objectionable scene in video [10]. In [4] porn video is detected based on motion features using HMM. Feature vectors are extracted by calculating the direction and magnitude after obtaining motion vectors from compressed MPEG video. These motion vectors are fed to Hidden Markov Model for classification of 6 actions. Skin color is one of important cues for pornographic video’s detection. More contributing towards classifying objectionable contents based on images can be found in [5]. They try to classify objectionable contents based on spectral audio signal features such as bandwidth (BW), frequency centric (FC), spectral flux (SF), spectral flatness (FT) and roll-off frequency (RO). In [6] they transformed

- Nilesh J. Uke, Assistant Professor, Department of Information Technology, Sinhgad College of Engineering, Pune, India and is currently pursuing PhD in Computer engineering in SRTM University, Nanded, India, PH-09850035805, E-mail: njuke.scoe@sinhgad.edu
- Dr. Ravindra C Thool, Professor and Head, Department of Information Technology, SGGGS College of Engineering and Technology, Nanded, India, PH- 02462-261264. E-mail: rcthol@sggs.ac.in

the color space, calculated Gaussian probability distribution and analyzed texture and noise to extract skin from a frame image, and then build a skin color mask. Finally algorithm analyzed the sensitive videos on the foundation of the current sensitive image recognition so as to understand them and shield the sensitive content.

NuDetective Forensic Tool [7] was developed in order to assist forensic examiners for conducting such analysis at the crime scene. Algorithm used for Image Analysis classifies each image as human nudity or not. The File Name Analysis verifies if files have suspicious names of child pornography. Erotic sound recognition for scene analysis based on Histogram Features and SVM is introduced for training and classification in [8]. Adaptive model to characterize human skin-color distributions for tracking human faces under different lighting conditions. The parameters of the model are adapted based on the maximum likelihood criterion. Detected skin regions are treated as blobs for extracting features representing pornographic images [9].

An array of image recognition techniques which includes face, nose, eye, mouth and upper body detected was used by [11] apart from skin detection. Work carried in [12] presents a technique for recognition of adult video using combination of skin detection features with motion information. Aim of using motion information is to select the appropriate color model that allows verifying pixels under different lighting conditions and other variations.

Objectionable video detection system presented in [13] uses hierarchical framework that consists three phases using multiple features in different temporal domains. Early detection is performed based on hash signatures prior to the download or play of a video in the first phase. Next phase consist of real-time detection which statistically estimates the degree to which the video may be objectionable by using single frame based features. Finally posterior detection is performed based on group-of-frame features reflecting the overall characteristics of the video. Fast and simple objectionable video classification scheme using temporal motion and color energy features (TMCEF) is proposed in [14]. Integrated system framework stated in TMCEF consists of key frame extraction, motion energy calculation, skin color energy calculation, and feature extraction based on statistical distribution metric using mean, standard deviation, and frequency analysis using discrete cosine transform (DCT). For verification the performance of these video-based temporal features, support vector machine (SVM) classifier is used.

3 WARFARE TOWARDS CHILD PORNOGRAPHY ON WEB

Protecting children from the red light districts of cyberspace parents could use the available resources on net. Handful of commercially developed products aimed at protecting children from inappropriate content on the Internet. Net Nanny uses real-time web content filtering and scans a website for pornography and other mature content and blocks them. Blocking pornography is important to prevent accidental exposure to inappropriate images.

Solid Oak Software's CYBERSitter brings internet filter software for protecting kids online. The program will only block the specific categories of content that are selected. Cyber Patrol, combat filter content, such as adult sites and unwanted computer programs, and to restrict kids' access to chat rooms and instant messaging, protecting them from cyber bullies and online predators. Limiting the time

your kids spend online is an added feature.

WebAllow is a parental and access control utility that restricts viewing all but selected websites. List of "approved" websites are predefined by the user. Even if a link is leading to another page on another domain is clicked, this software will block it if is not on the "approved" list.

Safe Eyes parental monitoring software falls in the category of Internet parental control software. Safe Eyes is a flexible set of tools that allows anyone to determine the type Internet usage that is appropriate. This monitoring tool includes content, program, time usage controls. Many parents wish to control the content their family views online, but the tools to date have been some time complex, inflexible, and costly. Blocking procedures followed by filter companies are awful, often capricious, and occasionally outright negligent. CYBERSitter, Cyber Patrol, and SurfWatch do not make their lists of blocked sites publicly available

4 PROPOSED SYSTEM FOR DETECTING PORNOGRAPHIC VIDEO

One of the gravest concerns of browning of obscene material on internet is to prohibit kids to get abused at later stage, intentionally or un-intentionally. We proposed new integrated computer vision approach to detect obscene and pornographic videos from internet, which will help parents to stop child pornography.

Adult and pornographic video contains large amount of movement information, contract changes from lenses from frequent zooming and tilting. These features can be used to detect obscene video. Erotic sound is another feature which is combined to get greater accuracy in detection process. In our proposed work we divided the entire system in three phases as shown in the Fig 1. Segmentation phase accept the input video stream and extract the key frames which are candidates for nude images detection. Amount of nudity will be detected from rapid moving object detection phase. Classification of objectionable video will be marked porn or non-porn depending upon the judgment criteria.

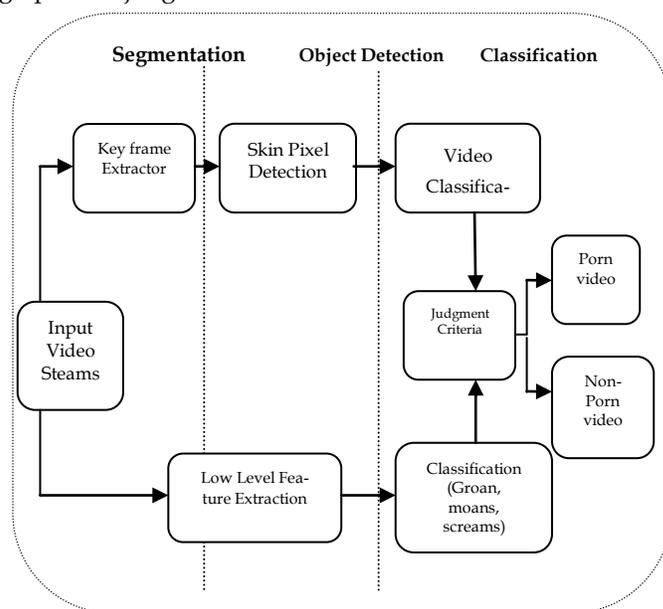


Fig.1 Block diagram

4.1 Segmentation and Skin Detection

If human skin color information is used efficiently for different lighting conditions then it can become an efficient tool for identifying facial and body features. Since it is invariant to rotation and partial occlusion, these features can be used for detection and localization. 2D histogram based segmentation will be used after key frame extraction for possible task of filtering adult content in the video.

4.2 Audio Based Classification

Detection of adult video will be carried on both visual interpretation and audio signal. Non-explicit sexual activity includes sound on an audio track, such as the kinds of groans, moans, and other sounds that to a reasonable person would imply sexual activity was taking place. Judgment criteria agent will have the patterns of audio signals for guessing such activities.

4.3 Video Based classification

Intel's OpenCV library is used to detect and classify video objects. Such pornographic video recognition process consists in extracting features that discriminate adult video and non-adult ones. Since uncovered skin distribution is of the almost importance for the detection of such videos, we take up the method to detect maximum skin pixels below abdomen. We know that sensitive body image has the larger area of bare skin than the normal image, calculating percentage of bare skin visibility makes up to judgment criteria.

5 CONCLUDING REMARKS AND PROGRESS OF WORK

Parents and children are using Internet based technology for teaching learning and other purposes. This work is coordination between computer vision techniques to solve a social problem of pornographic/adult video detection. Detecting such obscene video at browser level itself would be perfect to avoid intentional and un-intentional viewing of adult material by children in their teenage. Having established that existing research into tradition pornography is not enough to confirm the existence of child pornography. Therefore combination of computer vision techniques may develop one significant step towards for parents and assisting law enforcement agencies to locate evidences and prosecute possible criminals. We are interested in evaluating the performance of our system after implementation when tested with non-child pornography datasets consisting of legal pornography. System will be tested in the laboratory with online websites.

REFERENCES

[1] West, Caroline, "Pornography and Censorship", The Stanford Encyclopedia of Philosophy.
[2] Zhiqiang Wen; Yanhui Zhu; Zhaoyi Peng; , "Survey on Web Image Content-Based Filtering Technology," Information Science and Engineering (ICISE), 2009 1st International Conference on , vol., no., pp.1463-1466, 26-28 Dec. 2009

[3] D. A. Forsyth and M. M. Fleck. Automatic detection of human nudes. *IJCV*, 32(1):63-77, 1999
[4] Zhiyi Qu; Yanmin Liu; Ying Liu; Kang Jiu; Yong Chen; , "A Porn Video Detecting Method Based on Motion Features Using HMM," *Computational Intelligence and Design*, 2009. *ISCID '09. Second International Symposium on* , vol.2, no., pp.461-464, 12-14 Dec. 2009
[5] JaeDeok Lim; SeungWan Han; ByeongCheol Choi; ByungHo Chung; ChoelHoon Lee; , "Classifying of Objectionable Contents with Various Audio Signal Features," *Information Technology Convergence and Services (ITCS)*, 2010 2nd International Conference on , vol., no., pp.1-5, 11-13 Aug. 2010
[6] Yun Wang; Xiaoyu Wu; Lei Yang; , "Sensitive body image detection technology based on skin color and texture cues," *Image and Signal Processing (CISP)*, 2010 3rd International Congress on , vol.6, no., pp.2661-2664, 16-18 Oct. 2010
[7] de Castro Polastro, M.; da Silva Eleuterio, P.M.; , "NuDetective: A Forensic Tool to Help Combat Child Pornography through Automatic Nudity Detection," *Database and Expert Systems Applications (DEXA)*, 2010 Workshop on , vol., no., pp.349-353, Aug. 30 2010-Sept. 3 2010
[8] ZiQiang Shi; BoYang Gao; JiQing Han; Zhen Wu; , "Study of Objectionable Sound Recognition Based on Histogram Features and SVM," *Image and Signal Processing*, 2009. *CISP '09. 2nd International Congress on* , vol., no., pp.1-4, 17-19 Oct. 2009
[9] Yang, J., Lu, W., and Waibel, A. Skin-color modeling and adaptation. In Chin, R. and Pong, T.C., editors, *3rd Asian Conf. on Computer Vision*, volume 1352 of LNCS, pp: 687-694, 1998
[10] Lv Li, Yang Shutang, Objectionable Videos Detection Algorithm Based on Optical Flow, "Journal of Computer Engineering", Vol. 33, No.12 pp. 220-221, 2007
[11] Xinyu Xing, Yu-Li Liang, Hanqiang Cheng, Jianxun Dang, Sui Huang, "SafeVchat: Detecting Obscene Content and Misbehaving Users in Online Video Chat Services", *International World Wide Web Conference Committee (IW3C2)*, 2011
[12] Bouirouga, H.; Elfkihi Lrit, S.; Jilbab, A.; Aboutajdine, D.; , "Recognition of adult video by combining skin detection features with motion information," *Multimedia Computing and Systems (ICMCS)*, 2011 International Conference on , vol., no., pp.1-5, 7-9 April 2011
[13] Seungmin Lee; Woochul Shim; Sehun Kim; , "Hierarchical system for objectionable video detection," *Consumer Electronics, IEEE Transactions on* , vol.55, no.2, pp.677-684, May 2009
[14] Byeongcheol Choi; Seungwan Han; Byungho Chung; Jaechol Ryou; , "Design and performance evaluation of temporal motion and color energy features for objectionable video classification," *Advanced Information Management and Service (IMS)*, 2010 6th International Conference on , vol., no., pp.37-41, Nov. 30 2010-Dec. 2010

Double Sampling Product-Cum-Dual to Ratio Estimators for Finite Population Mean in Sample Surveys

B. K. Singh, Sanjib Choudhury

Abstract— This paper considers a class of product-cum- dual to ratio estimators for estimating finite population mean of the study variate in double sampling. The bias and mean square errors (MSEs) of the proposed estimator have been obtained in two different cases. The asymptotically optimum estimators (AOEs) of the class are identified along with their bias and MSEs. Theoretical and empirical studies have been done to demonstrate the efficiency of the proposed estimators over other estimators. An attempt has been made to find optimum sample sizes under a known fixed cost function.

Index Terms— Auxiliary variate; Finite population mean; Product estimator; Dual to ratio estimator; Double sampling; MSE; Optimum sample sizes, Efficiency.

1 INTRODUCTION

THE literature on survey sampling describes a great variety of techniques for using auxiliary information in order to obtain improved estimators for estimating some most common population parameters such as, population total, population mean, population proportion, population ratio, etc. More often, we are interested in the estimation of the mean of a certain characteristic of a finite population on the basis of a sample taken from the population following a specified sampling procedure.

The use of auxiliary information at the estimation stage appears to have started with the work of [22]. Cochran [22] used auxiliary information at estimation stage and proposed ratio estimator. Robson [5] and Murthy [13] envisaged product estimator and [6] used coefficient of variation of study variate. Motivated by [6], [3] utilized coefficient of variation of auxiliary variate. Srivenkataramana [20] was the first to propose dual to ratio estimator, [17] introduced dual variables for estimation of population parameters.

When the population mean \bar{X} of the auxiliary variable x is unknown before start of the survey, it is estimated from a preliminary large sample on which only the auxiliary characteristic x is observed. The value of \bar{X} in the estimator is then replaced by its estimate. After then a smaller second-phase sample of the variate of interest (study variate) y is then taken. This technique is known as double sampling or two-phase sampling. Neyman [9] was the first to give the concept of double sampling in connection with collecting information on the strata sizes in a stratified sampling, [7] considered ratio-product estimator in

double sampling, [18] proposed a generalized class of double sampling estimator based on ratio type estimator, [8] considered efficient estimator in double sample with subsample the non-respondents.

Consider a finite population $U = (u_1, u_2, \dots, u_N)$ of size N units, y and x are the study and auxiliary variates respectively. When the population mean \bar{X} of x is not known, a first-phase sample of size n_1 is drawn from the population on which only the x characteristic is measured in order to furnish a good estimate of \bar{X} . After then a second-phase sample of size n ($n < n_1$) is drawn on which both the variates y and x are measured.

The usual product estimator in double sampling is given as

$$\bar{y}_p^{(d)} = \bar{y} \frac{\bar{x}}{\bar{x}_1}$$

where \bar{x} and \bar{y} are the sample mean of x and y respectively based on the sample size n out of the population N units and

$\bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i$ denote the sample mean of x based on the first-

phase sample of the size n_1 .

Consider a transformation $x_i^\sigma = (N\bar{X} - nx_i)/(N-n)$, ($i=1, 2, 3, \dots, N$). Here $\bar{x}^\sigma = (N\bar{X} - n\bar{x})/(N-n)$ is unbiased estimator for \bar{X} and $Corr(\bar{y}, \bar{x}^\sigma) = -\rho$, where ρ is the correlation coefficient between y and x .

Using the transformation of x_i^σ , [20] obtained dual to ratio estimator as

- B. K. Singh, Department of Mathematics, North Eastern Regional Institute of Science and Technology, Nirjuli-791109, Arunachal Pradesh, India. E-mail: bksinghnerist@gmail.com
- Sanjib Choudhury (Corresponding Author), Department of Mathematics, North Eastern Regional Institute of Science and Technology, Nirjuli-791109, Arunachal Pradesh, India, E-mail: sanjibchy07@gmail.com

$$\bar{y}_R^* = \bar{y} \frac{\bar{x}^\sigma}{\bar{X}}$$

Utilizing the transformation $x_i^* = (n_1 \bar{x}_1 - n x_i) / (n_1 - n)$, ($i = 1, 2, \dots, N$), [12] has obtained dual to ratio estimator in double sampling as

$$\bar{y}_k^{(d)} = \bar{y} \frac{\bar{x}^*}{\bar{x}_1}$$

where $\bar{x}^* = (n_1 \bar{x}_1 - n \bar{x}) / (n_1 - n)$ be an unbiased estimator of \bar{X} and $Corr(\bar{y}, \bar{x}^*) = -\rho$.

In this paper, we have proposed an estimator of a linear combination of usual product estimator and dual to ratio estimator in double sampling. Numerical illustrations are given in the support of the present study.

2 THE PROPOSED ESTIMATOR

The proposed product-cum-dual to ratio estimators of population mean \bar{Y} in double sampling is

$$\bar{y}_{PDR}^{(d)} = \bar{y} \left[\alpha \frac{\bar{x}}{\bar{x}_1} + (1 - \alpha) \frac{\bar{x}^*}{\bar{x}_1} \right] \quad (1)$$

where α is a constant.

To obtain the bias (B) and mean square error (MSE) of $\bar{y}_{PDR}^{(d)}$, we write

$$e_0 = (\bar{y} - \bar{Y}) / \bar{Y}, \quad e_1 = (\bar{x} - \bar{X}) / \bar{X} \quad \text{and} \quad e_1' = (\bar{x}_1 - \bar{X}) / \bar{X}.$$

Expressing $\bar{y}_{PDR}^{(d)}$ in terms of e 's, we obtain

$$\begin{aligned} \bar{y}_{PDR}^{(d)} - \bar{Y} &= \bar{Y} (1 + e_0) \left[\alpha (1 + e_1) (1 + e_1')^{-1} \right. \\ &\quad \left. + (1 - \alpha) \left\{ (1 + g) - g (1 + e_1) (1 + e_1')^{-1} \right\} \right] \end{aligned} \quad (2)$$

where $g = n / (n_1 - n)$.

Assuming that the sample size is large enough so that $|e_1'| < 1$,

therefore $(1 + e_1')^{-1}$ is expandable.

Expanding the right hand side of (2), multiplying out and retaining terms of e 's up to the second degree, we obtain

$$\bar{y}_{PDR}^{(d)} - \bar{Y} \cong \bar{Y} \left\{ e_0 + \omega (e_1 - e_1' + e_1'^2 + e_0 e_1 - e_0 e_1' - e_1 e_1') \right\}, \quad (3)$$

where $\omega = (1 + g) \alpha - g$.

To obtain the bias and MSE of $\bar{y}_{PDR}^{(d)}$, the following notations are used hereafter:

$$C_y^2 = S_y^2 / \bar{Y}^2, \quad C_x^2 = S_x^2 / \bar{X}^2 \quad \text{and} \quad \rho = S_{xy} / (S_x S_y),$$

where C_y and C_x are the coefficient of variation of study variate y and auxiliary variate x respectively.

$S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2$ and $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2$ are the population variances of study variate y , auxiliary variate x respectively and

$S_{xy} = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})(x_i - \bar{X})$ is the covariance between y and x .

The following two cases will be discussed separately.

Case I: When the second phase sample of size n is a subsample of the first phase sample of size n_1 and

Case II: When the second phase sample of size n is drawn independently of the first phase sample of size n_1 , see [4].

3 BIAS, MSE AND OPTIMUM OF $\bar{y}_{PDR}^{(d)}$ IN CASE I

In Case I, we have

$$\left. \begin{aligned} E(e_0) &= E(e_1) = E(e_1') = 0, \\ E(e_0^2) &= \frac{1-f}{n} C_y^2, \quad E(e_1^2) = \frac{1-f}{n} C_x^2, \\ E(e_1'^2) &= \frac{1-f_1}{n_1} C_x^2, \quad E(e_0 e_1) = \frac{1-f}{n} C C_x^2, \\ E(e_0 e_1') &= \frac{1-f_1}{n_1} C C_x^2, \quad E(e_1 e_1') = \frac{1-f_1}{n_1} C_x^2, \end{aligned} \right\} \quad (4)$$

where $f = n/N$, $f_1 = n_1/N$ and $C = \rho C_y / C_x$.

Taking expectation on both the sides and using the results of (4) in (3), we get the bias of the estimator $\bar{y}_{PDR}^{(d)}$ to the first degree of approximation as

$$B \left\{ \bar{y}_{PDR}^{(d)} \right\}_I = \frac{1-f^*}{n} \bar{Y} C C_x^2 \omega, \quad (5)$$

where $f^* = n/n_1$.

The bias, $B \left\{ \bar{y}_{PDR}^{(d)} \right\}_I$ in (5) is 'zero' if $\alpha = n/n_1 (= f^*)$.

Thus the estimator $\bar{y}_{PDR}^{(d)}$ with $\alpha = n/n_1$ is almost unbiased.

To obtain MSE of proposed class of estimator, (3) can be written as

$$\bar{y}_{PDR}^{(d)} - \bar{Y} \cong \bar{Y} \left\{ e_0 + \omega (e_1 - e_1') \right\} \quad (6)$$

Squaring and taking expectation on both the sides of (6) and using the results of (4), we obtain the MSE of $\bar{y}_{PDR}^{(d)}$ up to the first degree of approximation as

$$M \left\{ \bar{y}_{PDR}^{(d)} \right\}_I = \bar{Y}^2 \left\{ \frac{1-f}{n} C_y^2 + \frac{1-f^*}{n} C_x^2 \omega (\omega + 2C) \right\} \quad (7)$$

Minimization of (7) with respect to α yields its optimum value at

$$\alpha = f^* - \frac{C}{1+g} = \alpha_{\text{opt}} \text{ (say).} \tag{8}$$

Substituting the value of α from (8) in (1) gives the 'asymptotically optimum estimator' (AOE) as

$$\left\{ \bar{y}_{PdR}^{(d_0)} \right\}_I = \bar{y} \left\{ \left(f^* - \frac{C}{1+g} \right) \frac{\bar{x}}{\bar{x}_1} + \frac{1+C}{1+g} \frac{\bar{x}^*}{\bar{x}_1} \right\}$$

Therefore, the resulting bias and MSE of $\bar{y}_{PdR}^{(d_0)}$ are

$$B \left\{ \bar{y}_{PdR}^{(d_0)} \right\}_I = -\frac{1-f^*}{n} \bar{Y} C^2 C_x^2$$

and

$$M \left\{ \bar{y}_{PdR}^{(d_0)} \right\}_I = \bar{Y}^2 C_y^2 \left(\frac{1-f}{n} - \frac{1-f^*}{n} \rho^2 \right) \text{ respectively.} \tag{9}$$

Equation (9) shows that MSE of the proposed estimator is the same as the MSE of the linear regression estimator in double sampling $\bar{y}_{dlr} = \bar{y} + b_{yx}(\bar{x}_1 - \bar{x})$, where b_{yx} is the sample regression coefficient of y on x .

Remark 1. For $\alpha = 1$, the estimator $\bar{y}_{PdR}^{(d)}$ in (1) reduces to the usual product estimator $\bar{y}_P^{(d)}$ in double sampling. The bias and MSE of $\bar{y}_P^{(d)}$ can be obtained by putting $\alpha = 1$ in (5) and (7) respectively as

$$B \left\{ \bar{y}_P^{(d)} \right\}_I = \bar{Y} \frac{1-f^*}{n} C C_x^2$$

and

$$M \left\{ \bar{y}_P^{(d)} \right\}_I = \bar{Y}^2 \left\{ \frac{1-f}{n} C_y^2 + \frac{1-f^*}{n} C_x^2 (1+2C) \right\} \tag{10}$$

Remark 2. For $\alpha = 0$, the estimator $\bar{y}_{PdR}^{(d)}$ in (1) reduces to dual to ratio estimator $\bar{y}_k^{(d)}$ in double sampling. The bias and MSE of $\bar{y}_k^{(d)}$ can be obtained by putting $\alpha = 0$ in (5) and (7) respectively as

$$B \left\{ \bar{y}_k^{(d)} \right\}_I = -\bar{Y} \frac{1-f^*}{n} g C C_x^2$$

and

$$M \left\{ \bar{y}_k^{(d)} \right\}_I = \bar{Y}^2 \left\{ \frac{1-f}{n} C_y^2 + \frac{1-f^*}{n} C_x^2 g (g-2C) \right\} \tag{11}$$

The variance of sample mean \bar{y} under SRSWOR sampling scheme is given by

$$V(\bar{y}) = \bar{Y}^2 \frac{1-f}{n} C_y^2 \tag{12}$$

4 EFFICIENCY COMPARISONS OF $\bar{y}_{PdR}^{(d)}$ AND $\bar{y}_{PdR}^{(d_0)}$ IN CASE I

In the following Section, we have presented the comparisons

of the proposed estimator with other estimators to investigate the ranges of the unknown parameter α for which the proposed estimator is better than others.

From (7), (10), (11) and (12) it is found that the proposed class of estimators $\bar{y}_{PdR}^{(d)}$ is better than:

i. the usual product estimator $\bar{y}_P^{(d)}$ in double sampling if

$$\text{either } 1 > \alpha \text{ and } -\left\{ 2C/(1+g) + N_1^* \right\} < \alpha,$$

$$\text{or } 1 < \alpha \text{ and } -\left\{ 2C/(1+g) + N_1^* \right\} > \alpha,$$

where $N_1^* = (n_1 - 2n)/n_1$.

ii. the dual to ratio estimator $\bar{y}_k^{(d)}$ in double sampling if

$$\text{either } 0 > \alpha \text{ and } 2\left\{ f^* - C/(1+g) \right\} < \alpha,$$

$$\text{or } 0 < \alpha \text{ and } 2\left\{ f^* - C/(1+g) \right\} > \alpha.$$

iii. the sample mean per unit estimator \bar{y} if

$$\text{either } f^* > \alpha \text{ and } \left\{ f^* - 2C/(1+g) \right\} < \alpha,$$

$$\text{or } f^* < \alpha \text{ and } \left\{ f^* - 2C/(1+g) \right\} > \alpha.$$

Also from (9), (10), (11) and (12), it is observed that the 'AOE' $\bar{y}_{PdR}^{(d_0)}$ is better than:

i. the usual product estimator $\bar{y}_P^{(d)}$ in double sampling, since

$$M \left\{ \bar{y}_P^{(d)} \right\}_I - M \left\{ \bar{y}_{PdR}^{(d_0)} \right\}_I = \bar{Y}^2 \frac{1-f^*}{n} C_x^2 (1+C)^2 > 0.$$

ii. the dual to ratio estimator $\bar{y}_k^{(d)}$ in double sampling, since

$$M \left\{ \bar{y}_k^{(d)} \right\}_I - M \left\{ \bar{y}_{PdR}^{(d_0)} \right\}_I = \bar{Y}^2 \frac{1-f^*}{n} C_x^2 (g-C)^2 > 0.$$

iii. the sample mean per unit estimator \bar{y} , since

$$V(\bar{y}) - M \left\{ \bar{y}_{PdR}^{(d_0)} \right\}_I = \bar{Y}^2 \frac{1-f}{n} C^2 C_x^2 > 0.$$

Now we state the following theorem:

Theorem 1. To the first degree of approximation, the proposed strategy $\bar{y}_{PdR}^{(d)}$ under optimal condition (8) is always more efficient than

$M \left\{ \bar{y}_P^{(d)} \right\}$, $M \left\{ \bar{y}_k^{(d)} \right\}$ and $V(\bar{y})$ and equally efficient with the linear regression estimator $M \left\{ \bar{y}_{dlr} \right\}$ in double sampling.

5 BIAS, MSE AND OPTIMUM OF $\bar{y}_{PdR}^{(d)}$ IN CASE II

In Case II, we have

$$\left. \begin{aligned} E(e_0) &= E(e_1) = E(e'_1) = 0, \\ E(e_0^2) &= \frac{1-f}{n} C_y^2, \quad E(e_1^2) = \frac{1-f}{n} C_x^2, \\ E(e_1'^2) &= \frac{1-f_1}{n_1} C_x^2, \quad E(e_0 e_1) = \frac{1-f}{n} C C_x^2, \\ E(e_0 e'_1) &= 0, \quad E(e_1 e'_1) = 0. \end{aligned} \right\} \quad (13)$$

Taking expectation in (3) and using the results of (13), we get the bias of $\bar{y}_{PdR}^{(d)}$ to the first degree of approximation as

$$B\{\bar{y}_{PdR}^{(d)}\}_{II} = \bar{Y} \omega C_x^2 \left(\frac{1-f_1}{n_1} + \frac{1-f}{n} C \right) \quad (14)$$

which is vanished when $\alpha = n/n_1 (= f^*)$. Thus, the estimators $\bar{y}_{PdR}^{(d)}$ with the value of $\alpha = n/n_1$ is almost unbiased.

Squaring and taking expectation in both the sides of (6) and using the results of (13), we obtain the MSE of $\bar{y}_{PdR}^{(d)}$ to the first degree of approximation as

$$M\{\bar{y}_{PdR}^{(d)}\} = \bar{Y}^2 \left\{ \frac{1-f}{n} C_y^2 + C_x^2 \omega \left(f^{**} \omega + 2 \frac{1-f}{n} C \right) \right\} \quad (15)$$

where $f^{**} = (1-f)/n + (1-f_1)/n_1$.

Minimization of (15) is obtained with optimum value of α as

$$\alpha = f^* - \left\{ \frac{1-f}{n} C \div f^{**} (1+g) \right\} = \alpha_{IIopt} \text{ (say)}. \quad (16)$$

Substituting the value of α from (16) in (1) gives the 'AOE' as

$$\left\{ \bar{y}_{PdR}^{(d)} \right\}_{II} = \left\{ \alpha_{IIopt} \frac{\bar{x}}{\bar{x}_1} + (1-\alpha_{IIopt}) \frac{\bar{x}^*}{\bar{x}_1} \right\}.$$

Thus, the resulting bias and MSE of 'AOE' $\bar{y}_{PdR}^{(d)}$ are respectively given as

$$B\left\{ \bar{y}_{PdR}^{(d)} \right\}_{II} = -\bar{Y} C C_x^2 \frac{1-f}{n} \left(\frac{1-f_1}{n_1} + \frac{1-f}{n} C \right) / f^{**}$$

and

$$M\left\{ \bar{y}_{PdR}^{(d)} \right\}_{II} = \bar{Y}^2 \left\{ \frac{1-f}{n} C_y^2 - \left(\frac{1-f}{n} \right)^2 C^2 C_x^2 / f^{**} \right\}.$$

For simplicity, we assume that the population size N is large enough as compared to the sample sizes n and n_1 so that the finite population correction (FPC) terms $1/N$ and $2/N$ are ignored.

Ignoring the FPC in (15), the MSE of $\bar{y}_{PdR}^{(d)}$ is reduces to

$$M\left\{ \bar{y}_{PdR}^{(d)} \right\}_{II} = \bar{Y}^2 \left[\frac{C_y^2}{n} + \omega C_x^2 \left\{ \left(\frac{1}{n} + \frac{1}{n_1} \right) \omega + 2 \frac{C}{n} \right\} \right] \quad (17)$$

which is minimized for

$$\alpha = f^* - c \mu = \alpha_{IIopt} \quad (18)$$

where $\mu = (n_1 - n)/(n_1 + n)$.

Substituting the value of α from (18) in (1), we obtain the 'AOE' as

$$\bar{y}_{PdR}^{(d_0^{**})} = \bar{y} \left[(f^* - C \mu) \frac{\bar{x}}{\bar{x}_1} + \left(\frac{1}{1+g} + C \mu \right) \left\{ (1+g) - g \frac{\bar{x}}{\bar{x}_1} \right\} \right]$$

Therefore, the resulting MSE of $\bar{y}_{PdR}^{(d_0^{**})}$ is

$$M\left\{ \bar{y}_{PdR}^{(d_0^{**})} \right\}_{II} = \frac{S_y^2}{n} (1 - \lambda \rho^2), \quad (19)$$

where $\lambda = n_1/(n_1 + n)$.

Remark 3. For $\alpha = 1$, the estimator $\bar{y}_{PdR}^{(d)}$ in (1) reduces to the usual product estimator $\bar{y}_p^{(d)}$ in double sampling. Thus putting $\alpha = 1$ in (17), we get the MSE of $\bar{y}_p^{(d)}$ to the first degree of approximation as

$$M\left\{ \bar{y}_p^{(d)} \right\}_{II} = \bar{Y}^2 \left[\frac{1}{n} \left\{ C_y^2 + C_x^2 (1+2C) \right\} + \frac{1}{n_1} C_x^2 \right]. \quad (20)$$

Remark 4. For $\alpha = 0$, the estimator $\bar{y}_{PdR}^{(d)}$ in (1) reduces to dual to ratio estimator $\bar{y}_k^{(d)}$ in double sampling. The MSE of $\bar{y}_k^{(d)}$ can be obtained by putting $\alpha = 0$ in (17) as

$$M\left\{ \bar{y}_k^{(d)} \right\}_{II} = \bar{Y}^2 \frac{1}{n} \left\{ C_y^2 - g C_x^2 \left(2C - \frac{g}{\lambda} \right) \right\} \quad (21)$$

Ignoring the FPC, the variance of \bar{y} under SRSWOR is given by

$$V(\bar{y})_{II} = \frac{1}{n} \bar{Y}^2 C_y^2 \quad (22)$$

and the MSE of the linear regression estimator $\bar{y}_{dR} = \bar{y} + b_{yx} (\bar{x}_1 - \bar{x})$ in double sampling is given by

$$M(\bar{y}_{dR})_{II} = \frac{S_y^2}{n} \left(1 - \frac{\rho^2}{1+g} \right). \quad (23)$$

6 EFFICIENCY COMPARISONS OF $\bar{y}_{PdR}^{(d)}$ AND $\bar{y}_{PdR}^{(d_0^{**})}$ IN CASE II

From (17), (20), (21) and (22) it is found that the proposed class of estimators $\bar{y}_{PdR}^{(d)}$ has better efficiency than

- i. the usual product estimator $\bar{y}_p^{(d)}$ in double sampling if either $1 > \alpha$ and $-(2C\mu + N_1^*) < \alpha$, or $1 < \alpha$ and $-(2C\mu + N_1^*) > \alpha$.
- ii. the dual to ratio estimator $\bar{y}_k^{(d)}$ in double sampling if

either $0 > \alpha$ and $2(f^* - \mu C) < \alpha$,

or $0 < \alpha$ and $2(f^* - \mu C) > \alpha$.

iii. the sample mean per unit estimator \bar{y} if

either $f^* > \alpha$ and $(f^* - 2C\mu) < \alpha$,

or $f^* < \alpha$ and $(f^* - 2C\mu) > \alpha$.

Also from (19), (20), (21), (22) and (23), it is established that

the 'AOE' $\bar{y}_{PdR}^{(d_0^{**})}$ is better than

i. the usual product estimator $\bar{y}_P^{(d)}$ in double sampling, since

$$M \left\{ \bar{y}_P^{(d)} \right\}_{II} - M \left\{ \bar{y}_{PdR}^{(d_0^{**})} \right\}_{II} = \bar{Y}^2 C_x^2 \lambda (C + 1/\lambda)^2 > 0.$$

ii. the dual to ratio estimator $\bar{y}_k^{(d)}$ in double sampling, since

$$M \left\{ \bar{y}_k^{(d)} \right\}_{II} - M \left\{ \bar{y}_{PdR}^{(d_0^{**})} \right\}_{II} = \bar{Y}^2 C_x^2 n \lambda \left(\frac{C}{n} - \frac{1}{\mu n_1} \right)^2 > 0.$$

iii. the sample mean per unit estimator \bar{y} , since

$$V(\bar{y})_{II} - M \left\{ \bar{y}_{PdR}^{(d_0^{**})} \right\}_{II} = \bar{Y}^2 C^2 C_x^2 \frac{\lambda}{n} > 0.$$

iv. the linear regression estimator \bar{y}_{dlr} in double sampling, since

$$M(\bar{y}_{dlr})_{II} - M \left\{ \bar{y}_{PdR}^{(d_0^{**})} \right\}_{II} = S_y^2 \rho^2 \frac{n\lambda}{n_1^2} > 0.$$

Remark 5. The MSE of the 'AOE' $\bar{y}_{PdR}^{(d_0^{**})}$ is always less than that of

\bar{y}_{dlr} . In addition to it, the 'AOE' $\bar{y}_{PdR}^{(d_0^{**})}$ in Case II is more efficient than the 'AOE' $\bar{y}_{PdR}^{(d)}$ in Case I.

7 COST ASPECT

The different estimators reported in this paper have so far been compared with respect to their mean square error. In practical applications, the cost aspect should also be taken into account. In literature, therefore, convention is to fix the total cost of the survey and then to find optimum sizes of preliminary and final samples so that the MSE of the estimators are minimized. In most of the practical situations, total cost is a linear function of samples selected at first and second phases.

In this section, we shall consider the cost of the survey and find the optimum sizes of the preliminary and second-phase samples in Case I and Case II separately.

Case I: In this case, Let us consider a cost function is given by

$$c = c_1 n + c_2 n_1 \tag{24}$$

where c = total cost of the survey.

c_1 = cost per unit of collecting information on y and

c_2 = cost per unit of collecting information on x .

Ignoring FPC, the MSE expression of 'AOE' $\bar{y}_{PdR}^{(d_0)}$ of (9) is given as

$$M \left\{ \bar{y}_{PdR}^{(d_0)} \right\} = \frac{V_1}{n} + \frac{V_2}{n_1} \tag{25}$$

where $V_1 = S_y^2 (1 - \rho^2)$ and $V_2 = S_x^2 \rho^2$.

The optimum values (sizes) of n and n_1 for fixed cost c , which minimizes the MSE of (25) are given by

$$\begin{aligned} n_{opt} &= c \sqrt{V_1/c_1} / \left(\sqrt{V_1 c_1} + \sqrt{V_2 c_2} \right) \\ &= c \sqrt{(1 - \rho^2)/c_1} / \left(\sqrt{(1 - \rho^2) c_1} + \rho \sqrt{c_2} \right) \text{ and} \\ n_{1opt} &= c \sqrt{V_2/c_2} / \left(\sqrt{V_1 c_1} + \sqrt{V_2 c_2} \right) \\ &= c \rho \sqrt{1/c_2} / \left(\sqrt{(1 - \rho^2) c_1} + \rho \sqrt{c_2} \right). \end{aligned}$$

Therefore, the MSE of $\bar{y}_{PdR}^{(d_0)}$ corresponding to optimal double sampling estimator is

$$M_{opt} \left\{ \bar{y}_{PdR}^{(d_0)} \right\} = (S_y^2/c) \left\{ \sqrt{(1 - \rho^2) c_1} + \rho \sqrt{c_2} \right\}^2 \tag{26}$$

In case we do not use any auxiliary variate then the cost function is of the form $c_0 = n c_1$, where $c_0 (= c)$ and c_1 are the total cost and cost per unit of collecting information on the study variate y respectively. The minimum variance of the sample mean \bar{y} for a given fixed cost c_0 in case of large population is given by

$$V_{opt}(\bar{y}) = \frac{c_1}{c_0} S_y^2 \tag{27}$$

From (26) and (27), the proposed double sampling strategy would be profitable than sample mean \bar{y} if

$$M_{opt} \left\{ \bar{y}_{PdR}^{(d_0)} \right\} < V_{opt}(\bar{y})$$

or equivalently,

$$\frac{c_2}{c_1} < \left(1 - \sqrt{1 - \rho^2} \right)^2 / \rho^2. \tag{28}$$

Case II: In this case we assume that x is measured on $n^{**} = n + n_1$ units and y on n units. Following [19], we shall consider a simple cost function as

$$c = c_1 n + c_2 n^{**}$$

where c_1 and c_2 denotes the cost per unit observing y and x values respectively.

The MSE of $\bar{y}_{PdR}^{(d_0^{**})}$ at (19) can be written as

$$M \left\{ \bar{y}_{PdR}^{(d_0^{**})} \right\} = \frac{V_1}{n} + \frac{V_2}{n^{**}} \tag{29}$$

The optimum values (sizes) of n and n^{**} for fixed cost c ,

which minimizes the MSE of (29) are given by

$$n_{opt} = c\sqrt{V_1/c_1} / \left(\sqrt{V_1c_1} + \sqrt{V_2c_2^*} \right)$$

$$= c\sqrt{(1-\rho^2)/c_1} / \left(\sqrt{(1-\rho^2)c_1} + \rho\sqrt{c_2^*} \right) \text{ and}$$

$$n_{opt}^{**} = c\sqrt{V_2/c_2^*} / \left(\sqrt{V_1c_1} + \sqrt{V_2c_2^*} \right)$$

$$= c\rho\sqrt{1/c_2^*} / \left(\sqrt{(1-\rho^2)c_1} + \rho\sqrt{c_2^*} \right)$$

The MSE of $\bar{y}_{PdR}^{(d_0^{**})}$ corresponding to optimal double sampling estimator is

$$M_{opt} \left\{ \bar{y}_{PdR}^{(d_0^{**})} \right\} = (S_y^2/c) \left\{ \sqrt{(1-\rho^2)c_1} + \rho\sqrt{c_2^*} \right\}^2 \quad (30)$$

From (27) and (30), it is obtained that the double sampling estimator $\bar{y}_{PdR}^{(d_0^{**})}$ yields less MSE than that of sample mean \bar{y} for the same fixed cost if

$$M_{opt} \left\{ \bar{y}_{PdR}^{(d_0^{**})} \right\} < V_{opt}(\bar{y})$$

or equivalently,

$$\frac{c_2^*}{c_1} < \left(1 - \sqrt{1-\rho^2} \right)^2 / \rho^2 \quad (31)$$

8 EMPIRICAL STUDY

To analyze the performance of the proposed estimators in comparison to other estimators, six natural population data sets are being considered. The sources of the population, the nature of the variates y and x and the values of the various parameters are given as follows

Population I- Source: [14], p. 228

x : fixed capital,
 y : output,

$$N = 80, n = 10, n_1 = 30, \bar{Y} = 5182.64, C_y = 0.3542,$$

$$C_x = 7507, \rho = 0.9413$$

Population II- Source: [14], p. 228

x : number of workers,
 y : output,

$$N = 80, n = 10, n_1 = 30, \bar{Y} = 5182.64, C_y = 0.3542,$$

$$C_x = 0.9484, \rho = 0.9150$$

Population III- Source: [2]

x : number of agricultural labourers for 1961,
 y : number of agricultural labourers for 1971,

$$N = 278, n = 30, n_1 = 70, \bar{Y} = 39.0680, C_y = 1.4451,$$

$$C_x = 1.6198, \rho = 0.7213$$

Population IV- Source: [15], p. 282

x : chlorine percentage,
 y : log of leaf burn in sacs.

$$N = 30, n = 4, n_1 = 12, \bar{Y} = 0.6860, C_y = 0.4803,$$

$$C_x = 0.7493, \rho = -0.4996$$

Population V- Source: [12], P. 324

x : population of village,
 y : number of cultivators in the village.

$$N = 487, n = 20, n_1 = 95, \bar{Y} = 449.846, C_y = 0.8871,$$

$$C_x = 0.7696, \rho = 0.881815.$$

Population VI- Source: [1], P. 47

x : initial white blood cell count,
 y : survival time leukemia patient.

$$N = 20, n = 4, n_1 = 8, C_y = 0.2017, C_x = 0.1502,$$

$$\rho = -0.4074.$$

To reflect the gain in the efficiency of the proposed class of estimators $\bar{y}_{PdR}^{(d)}$ over the conventional estimators (\bar{y} and $\bar{y}_p^{(d)}$)

and dual to ratio estimator $\bar{y}_k^{(d)}$ in double sampling, the effective ranges and optimum values of α are presented in Table 1 with respect to the above population data sets.

TABLE 1 EFFECTIVE RANGES AND OPTIMUM VALUES OF α OF $\bar{y}_{PdR}^{(d)}$

Popula- tion data sets	Ranges of α under which the proposed class of estimators is better than			Optimum values of α
	\bar{y}	$\bar{y}_p^{(d)}$	$\bar{y}_k^{(d)}$	
For Case I				
I	(-0.26, 0.33)	(-0.93, 1.00)	(0.00, 0.07)	α_{Iopt} 0.372
II	(0.12, 0.33)	(-0.79, 1.00)	(0.00, 0.21)	0.1055
III	(-0.31, 0.43)	(-0.88, 1.00)	(0.00, 0.12)	0.0609
IV	(0.33, 0.76)	(0.09, 1.00)	(0.00, 1.09)	0.5468
V	(-1.39, 0.21)	(-2.18, 1.00)	(-1.18, 0.00)	-0.5919
VI	(0.50, 1.05)	(0.55, 1.00)	(0.00, 1.55)	0.7735
For Case II				
I	(-0.11, 0.33)	(-0.78, 1.00)	(0.00, 0.22)	α_{IIopt}^* 0.1113
II	(-0.01, 0.33)	(-0.68, 1.00)	(0.00, 0.32)	0.1625
III	(-0.09, 0.43)	(-0.66, 1.00)	(0.00, 0.34)	0.1712
IV	(0.33, 0.65)	(-0.01, 1.00)	(0.00, 0.99)	0.4935
V	(-1.12, 0.21)	(-1.90, 1.00)	(-0.90, 0.00)	-0.4524
VI	(0.50, 0.86)	(0.36, 1.00)	(0.00, 1.36)	0.6823

To observe the relative performance of different estimators of \bar{Y} , we have computed the percentage relative efficiencies (PREs) of the proposed estimator, conventional estimators (i.e. \bar{y} and $\bar{y}_p^{(d)}$) and dual to ratio estimator $\bar{y}_k^{(d)}$ in double sampling with respect to usual unbiased estimator \bar{y} by using the formula

$$PRE(t, \bar{y}) = \frac{M(\bar{y})}{M(t)} \times 100,$$

where $t = \bar{y}, \bar{y}_p^{(d)}, \bar{y}_k^{(d)}$ and $\bar{y}_{PdR}^{(d)}$ or $\left(\bar{y}_{PdR}^{(d_0)} \text{ and } \bar{y}_{PdR}^{(d_0^{**})} \right)$.

and the findings are presented in Table 2.

TABLE 2 PERCENTAGE RELATIVE EFFICIENCIES WITH RESPECT TO \bar{y}

Estimators	\bar{y}	$\bar{y}_P^{(d)}$				$\bar{y}_{PdR}^{(d)}$ or $\bar{y}_{PdR}^{(d_0)}$ or $\bar{y}_{PdR}^{(d_0^{**})}$	
		Case I	Case II	Case I	Case II	Case I	Case II
I	100.00	*	*	297.97	199.08	307.77	298.09
II	100.00	*	*	200.42	106.40	276.16	268.76
III	100.00	*	*	147.96	125.49	149.98	157.28
IV	100.00	*	*	*	*	123.76	123.83
V	100.00	*	*	141.21	152.26	277.92	279.61
VI	100.00	103.38	*	*	*	111.56	112.44

*Data not applicable and percentage relative efficiency less than 100%.

9 CONCLUSIONS

Section 7 provides the optimum sample sizes under a known fixed cost function for Case I and Case II. We have also obtained conditions (referred in (28) and (31)) under which the proposed class of estimators $\bar{y}_{PdR}^{(d)}$ would be profitable in comparison to single-phase sampling for fixed cost of survey.

From Table 1, it is seen that the proposed class of estimators $\bar{y}_{PdR}^{(d)}$ is more efficient over the conventional estimators $(\bar{y} \text{ and } \bar{y}_P^{(d)})$ and dual to ratio estimator $(\bar{y}_k^{(d)})$ in double sampling for both the cases under the effective ranges of α as far as the MSE criterion is concerned. It is also observed from Table 1 that there is a scope for choosing α to obtain better estimators than \bar{y} , $\bar{y}_P^{(d)}$ and $\bar{y}_k^{(d)}$.

Table 2 shows that there is a considerable gain in efficiency by using proposed class of estimators $\bar{y}_{PdR}^{(d)}$ or $\left\{ \bar{y}_{PdR}^{(d_0)} \right\}$ and $\left\{ \bar{y}_{PdR}^{(d_0^{**})} \right\}$ over the conventional estimators $(\bar{y} \text{ and } \bar{y}_P^{(d)})$ and dual to ratio $(\bar{y}_k^{(d)})$ estimator in double sampling with respect to the population data sets in both the cases. This shows that even if the scalar α deviates from its optimum values $(\alpha_{Iopt} \text{ and } \alpha_{IIopt}^*)$, the proposed class of estimators $\bar{y}_{PdR}^{(d)}$ will yield better estimates than \bar{y} , $\bar{y}_P^{(d)}$ and $\bar{y}_k^{(d)}$. It is further observed that the estimator $\bar{y}_{PdR}^{(d_0)}$ is more efficient than $\bar{y}_{PdR}^{(d)}$, except for the data sets of population I and II, where the estimator $\bar{y}_{PdR}^{(d_0)}$ is slightly better than $\bar{y}_{PdR}^{(d)}$. Thus, the use of the proposed class of estimators is preferable over others.

REFERENCES

- [1] A. J. Dobson, *An Introduction to Generalized Linear Models*, 1st Edition, New York: Chapman & Hall, 1990.
- [2] A. K. Das, "Contribution to the theory of sampling strategies based on auxiliary information," Unpublished Ph. D. Thesis, Bidhan Chandra Krishi Vishwavidyalaya, India, 1988.
- [3] B. V. S. Sisodia and V. K. Dwivedi, "A modified ratio estimator using coefficient of variation of auxiliary variable," *Jour. Ind. Soc. Agril. Statist.*, vol. 33, pp. 13-18, 1981.
- [4] C. Bose, "Note on the sampling error in the method of double sampling," *Sankhyā*, vol. 6, pp. 329-330, 1943.
- [5] D. S. Robson, "Applications of multivariate polykeys to the theory of unbiased ratio-type estimation," *Journal of the American Statistical Association*, vol. 52, pp. 511-522, 1957.
- [6] D. T. Searls, "The utilization of known coefficient of variation in the estimation procedure," *Journal of American Statistical Association*, vol. 59, pp. 1125-26, 1964.
- [7] H. P. Singh and M. R. Espejo, "Double sampling ratio-product estimator of a finite population mean in sampling surveys," *Journal of Applied Statistics*, vol. 34, no. 1, pp. 71-85, 2007.
- [8] H. P. Singh and S. Kumar, "Improved estimation of population mean under two-phase sampling with subsampling the non-respondents," *Journal of Statistical Planning and Inference*, vol. 140, no. 9, pp. 2536-2550, 2010.
- [9] J. Neyman, "Contribution to the theory of sampling human populations," *Journal of the American Statistical Association*, vol. 33, pp. 101-116, 1938.
- [10] L. Chand, "Some ratio-type estimators based on two or more auxiliary variables," PhD Thesis, Ames, Iowa State University, 1975.
- [11] M. A. Hidiroglou and C. E. Sarndal, "Use of auxiliary information for two-phase sampling," *Survey Methodology*, vol. 24, pp. 11-20, 1998.
- [12] M. Kumar and S. Bahl, "Class of dual to ratio estimators for double sampling," *Statistical Papers*, vol. 47, pp. 319-326, 2006.
- [13] M. N. Murthy, "Product method of estimation," *Sankhyā A*, vol. 26, pp. 69-74, 1964.
- [14] M. N. Murthy, *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta: India, 1967.
- [15] R. G. D. Steel and J. H. Torrie, *Principles and Procedures of Statistics*, New York: McGraw-Hill, 1960.
- [16] R. R. Sitter, "Variance estimation for the regression estimator in two-phase sampling," *Journal of the American Statistical Association*, vol. 92, no. 438, pp. 780-787, 1997.
- [17] S. Bandyopadhyay, "Improved ratio and product estimators," *Sankhyā Series C*, vol. 42, no 2, pp. 45-49, 1980.
- [18] S. Bhushan, A. Pandey and S. Katara, "A class of estimators in double sampling using two auxiliary variables," *Journal of Reliability and Statistical Studies*, vol. 1, Issue 1, pp. 67-73, 2008.
- [19] S. K. Srivastava, "A two-phase sampling estimator in sample surveys," *Australian Journal of Statistics*, vol. 2, no. 1 pp. 23-27, 1970.
- [20] T. Srivenkataramana, "A dual to ratio estimator in sample surveys," *Biometrika*, vol. 67, no. 1, pp. 199-204, 1980.
- [21] W. G. Cochran, *Sampling Technique*, 2nd Edition, Wiley: New York, 1963.
- [22] W. G. Cochran, "The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce," *The Journal of Agricultural Science*, vol. 30, no. 2, pp. 262-275, 1940.

Design and Implementation of Environmental Teaching System of University

Shuangchun Yang, Yi Pan, Jinhui Zhang, Xiaoyang Geng, Xiaozhen Wang
Liaoning Shihua University, Fushun, China

Abstract— the traditional education model has many shortcomings of low teaching efficiency, poor quality of education in China. By the promotion of continuous improvement of the computer application technology, I innovative Environmental Teaching System (ETS) was put forward. In this paper, ETS teaching system consists of three subsystems: Environmental Lecture sub-System (ELsS), Environmental Assignment sub-System (EAsS), and Environmental Test sub-System (ETsS). The ETS also have five functions, which are Environmental Classroom Assessment, Environmental Classroom Records, Email-Based Assignment Management Model, Environmental Correcting Examination Paper and Online Examination Records. ETS has many advantages than traditional education model in the teaching process, such as convenient and efficient, data intact, saving resources and so on.

Index Terms— Environmental, Paperless, Teaching System, Teaching reform, Design, online examination, education model.

1 INTRODUCTION

With continuous development of the information technology, computers and networks are widely used in peoples' lives and working[1]. More and more people use digital tools to transfer and store information rather than paper. For instance, most universities are currently conducting multimedia teaching reform with the computer technology to replace the traditional blackboard chalk[2,3]. In the traditional teaching process, the teacher has many things to do like roll call records, send and receive homework, so the paper is needed. This process is very exhausting. There are many problems to be solved, such as, the recording archiving and long-term preservation.

At present, the development of paperless examination software is also emerging in the domestic and international[4,5,6]. But functions of the system are different each other because of different developers. Generally, this can be divided into two categories:

One is the established constant software for the examination of a particular discipline. This software can't be directly used in other examination. For example, in Computer Rank Examination and Traffic Laws Examination, test questions can't be modified, this software don't have universal applicability.

The other is a big lack of software programming, most software is programmed by software engineers. These engineers usually don't understand the process of teaching. Examination is the important way to evaluate the teaching effectiveness. If you don't know how to evaluate teaching effectiveness, then you can't serve the teaching well [3].

This paper presents a new teaching system- Environmental Teaching System (ETS). ETS makes digital technology be applied to the teaching and achieve the paperless process from the class to the examination. Multimedia teaching applied in universities is a part of ETS at present. In this paper, we will detail the design and implementation of ETS based on the actual situation of Liaoning Shihua University.

2 THE PROPOSITION OF ETS

Liaoning Shihua University is founded in Dalian, in 1950. It

is the first petroleum school in China. Its major specialty are Petroleum, Machinery and Information, at present, there are 23,430 students in the school. The fundamental teaching conditions in this school are computer and network. The application of computers in daily office work is very common. For example, if the students' affairs department has any notice, it can send electronic document by email to every college. Then every college will print a paper document to students. With the help of Internet, the student selecting course and query achievement are more convenient than before (as shown as Fig.1). The area of Liaoning shihua University are 1906 acres, it consists of main school area, professional technology college and Yingkou Campus. In the 2011 ~ 2012 school year, the curriculum of the author is as follows, on Monday, the class is Instrumental Analysis in the professional technology college and the principle of Food Engineering in main school area. On Wednesday, the class is Inorganic Chemistry in the professional technology college and Physical Pollution Control in main school area. For example, the amounts of the exercise book of Physical Pollution Control for two classes are 60, and the amounts of the exercise book of Inorganic Chemistry are 61. In the form of a single paper is not conducive to save. If we allow students to complete after class, students may copy the homework. Before class, teachers need to prepare textbooks, reference books, teaching manuals, portable laptop, even a pile of exercise books. This will virtually increase the burden of teachers. The idea of ETS is from this teaching reality.



Fig.1.web of student selecting course

3 THE REALIZATION OF ETS

As demonstrated in this document, the numbering for sections upper case Arabic numerals, then upper case Arabic numerals, separated by periods. Initial paragraphs after the section title are not indented. Only the initial, introductory paragraph has a drop cap.

In this paper, the proposed system framework(as shown as Fig.2) is shown below. ETS is consisted of three subsystems, Environmental Lecture sub-System(ELsS), Environmental Assignment sub-System(EAsS), and Environmental Test sub-System(ETsS).

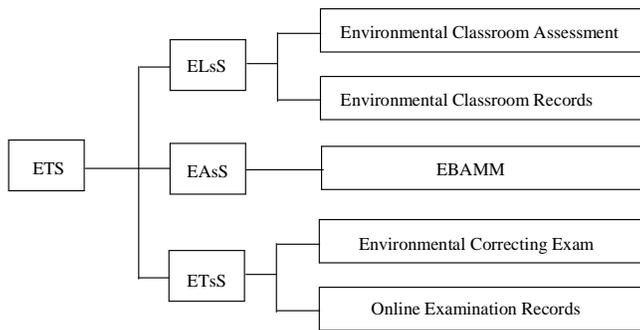


Fig.2.Framework of ETS

3.1 Environmental Lecture sub-System(ELsS)

ELsS includes Environmental classroom assessment and no paper classroom records.

1) Environmental classroom assessment

In the traditional classroom, teachers check the attendance rate by the roll call. When there are so many people, such as University Physics and basic course, the roll call will take up the large amount of teaching time. Environmental classroom assessment uses computer image recognition to complete the assessment work before class, it makes more teaching time available than before. We can install a fingerprint collector in each of the door of the classroom, students can complete sign before the class. Even the number of students is too much; the teacher can know instant students attendance by the fingerprint identification signed. When they enter the school, the school shall collect students' fingerprints to prevent signing for somebody else, and establish a database of students for ETS as shown in Figure 3:

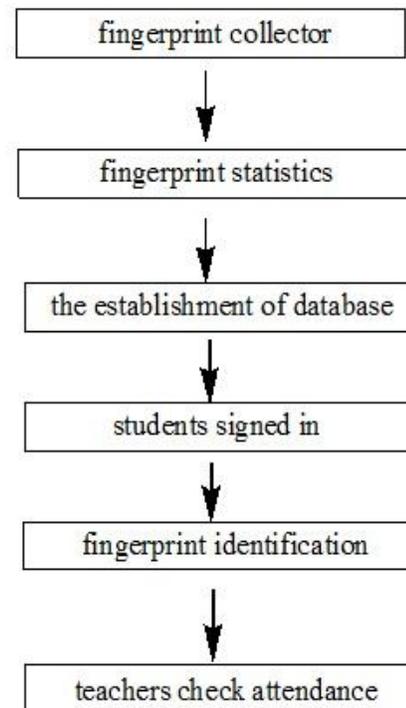


Fig.3 Framework of environmental classroom assessment

2) Environmental classroom records

Environmental classroom records is no longer follow the traditional teaching model which teacher lectures on the podium and students records by pen and paper. Students can download the PPT by the computer before class. If students don't understand the contents of the PPT, they can mark directly on it[7,8,9]. Students can add the focus of the content on the courseware. There may be many practical problems in the process of no paper classroom records, because we have no experience. Such as the intellectual property rights of the PPT, for this, we can take the following measures. The teacher can set password to the courseware, then tell students. And students need to land by their student id before download the important courseware. So we can guarantee the safety of the intellectual property right.

3.2 Environmental Assignment sub-System (EAsS)

In the traditional operation mode, teachers correct though the page; sometimes they need stand with scribbled pain. In every school year, the school leaders will spot check occasional of teachers work situation. So teachers need to save assignments to prepare for the examination in next semester. EAsS is more convenient than traditional operation mode, because a series of work about the operation will finish on the Web. For example, teachers can use the Amendment function of "Word" software to mark assignments. The marked content will be shown in red, which is very eye-catching. Teachers can use the Annotation function to write reviews of operations to provide students with specific comments and suggestions. The annotated part will be shown in red too. Then teachers can return marked and annotated assignments to students by the Internet. All records are preserved in the "Operation Database".

Teachers can correct assignments in any place with Internet access; they don't have to carry the operation of this exchange between the office and home. So teachers are very convenience by EPAsS and their work efficiency will be improved significantly^[10]. We will create the Email-Based Assignment Management Module(EBAMM). The entire process includes that Authentication, Students enrolling, Operation inputting, Assigning homework, Receiving operation, Grading assignment, Publication of the results and assignment statistics. As shown in figure 4:

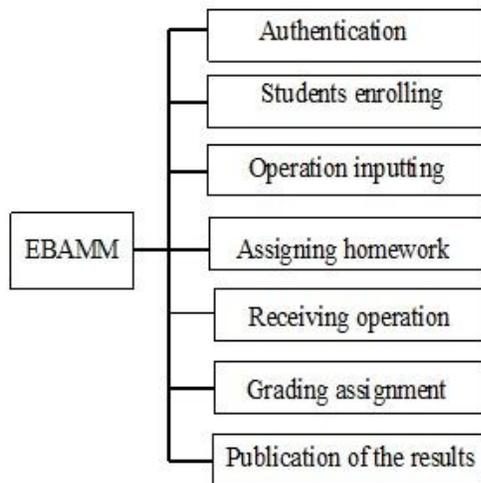


Fig.4 Framework of Email-Based Assignment Management Module

- (1) Authentication. EBAMM^[11,12] requires only the registered users to use the function of system. So, teachers can found account, set password, fill the teaching work Email and other personal information before use the system.
- (2) Students enrolling. Students shall enroll in the system before use Email to submit assignments to reduce burden of teachers. Students send personal information (such as student id, name Email address, etc.) to teachers according to the regulation of the format, teachers receive Email by the system; students' information will be filtered and wrote by the system to realize the automatic input of students' information.
- (3) Operation inputting. Teachers input homework to the system and establish homework library before assign the homework. The homework can be formed by group volume with the help of the external database system and can also be introduced directly by external files. At the same time, the answer of homework will be saved in the system database.
- (4) Assigning homework. Teachers first choose the assignment questionnaire; make sure contents of homework, then set the due time. The system will write the information to database with the request of teachers, and take the questionnaire out of the database, then send to students in the form of Email attachments.
- (5) Receiving operation. Students shall name the task accessories by the specified way before submit operations, in order to teachers can save assignments orderly.
- (6) Grading assignment. The system will chose all the home-

- work that not given the result from the database, then teachers can correct the homework documents by the computer. After reviewing the homework, teachers will give achievement and comment, and write into database.
- (7) Publication of the results. The system will chose all the homework that have given the result but not release (default is "not release") from the database, and send the answer and excellent students' assignment to every student by Email to contrast.
- (8) Assignment statistics. The system has the function that automatic statistics, which can give detailed statistical reports. When the submission deadline arrives, the system will list the statistics of students have not submitted the assignment, and generate automatically reminders email, then send to the appropriate mailbox.

Problems of no paper assignment have been solved detailed by EBAMM. EBAMM system not only saves resources and saves time and effort, but also makes for saving assignment. The educational administration departments can retrieval of database through the Internet to check the situation of grading assignment.

3.3 Environmental Test sub-System (ETsS)

ETsS consists of no paper correcting examination paper and online examination records.

In the traditional examination process, the school requires a lot of manpower and material resources^[13]. For example, the school needs to print a large number of papers, needs invigilator during the examination and requires teachers marking papers after exam. It is a waste of paper resources and time. Every academy need to save examination papers to prepare for educational administration department to check up on it at the beginning of each semester. The process of ETsS is shown as follows, Establishment of test database – Select test questions - Test online - Marking subjective questions online - Teachers marking objective questions by computers - Submit - Uniform marking the papers and Store the results in the database - To save.

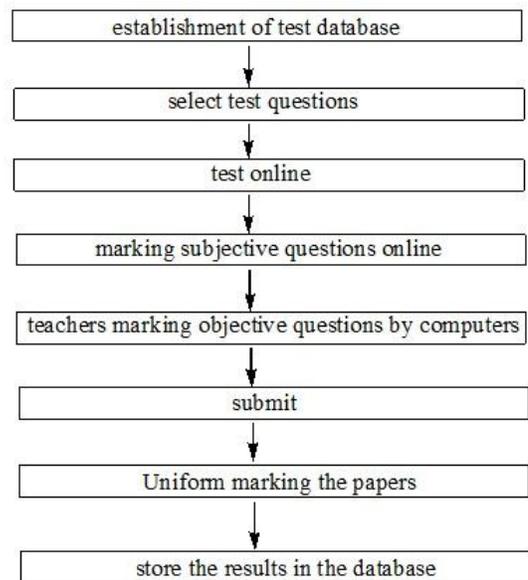


Fig.5 Framework of Environmental Test sub-System

The environmental correcting examination paper system can use the machine-readable to read over objective questions that greatly reduce the workload of marking. The machine-readable is very simple. All universities have the corresponding conditions. The system shall have the function which is Online Subjective Marking. Teachers can mark the examination paper online. The same questions can also be marked by many reviewers online, respectively. Scores derived from the weighted average. So the result will be more objective than before, and the quality of marking examination paper will be improved^[14].

The problem we need to note on online examination records is cheating. The examination records of ETS makes students test by computers. Schools need to install the device identification code to the existing computer room to ensure that students themselves take exam. In the examination room, real-time surveillance cameras have been installed on each test computer. Candidate login screen will appear after fingerprint identification through the computer. Then candidates can input the information into the examination system for examination. Some problems in the examination process such as window control, taskbar control, hotkey shielding and some messages of the application have been resolved already^[15].

Whether the online examination system can really play a role in practical applications, it requires a large number of test questions. If the number of questions is small, students may be a simple answer back, which deviate from the original intention of the network test. If there are a sufficient number of questions, students will turn to understanding the basal knowledge. Students will find their blind spot of knowledge in the process of study, and keep abreast of new knowledge. It will greatly promote the development of teaching. To this end we should establish our own Test Bank. In the beginning of each semester, there are some groups of teachers that teach the same course, they will set questions, then proofread questions each other in the same group. We will publish each teacher's name on the landing page of the examination system. On the one hand students will understand the workload of teachers, the other hand, if there is some problems of questions, teachers will be informed and correct errors in time. And teachers will enhance their sense of responsibility. So the test database will run long-term with health.

4 THE STEPS OF ETS ENVIRONMENTAL TEACHING SYSTEM IMPLEMENTATION

ETS actual implementation of paperless education system needs to solve some problems.

4.1 Technical issues

ETS system will break the traditional mode of education; it will be fully able to carry out based on existing technology. The whole ETS teaching paperless system will not be off computers and networks from paperless system in the classroom to the end of the environmental examination system, so this requires the school should invest a lot of money to improve related facilities, such as multimedia resources and device identification code.

4.2 Pre-training

Currently, there is no one university has fully implemented a environmental paperless education system, of course, which requires pre-service training of college teachers before the implementation of ETS teaching system.

4.3 The implementation from the part to the overall

Take Liaoning Shihua University for example, the school has three campuses and this is a new measure, some problems need to explore, so the next step is to test in a campus.

5 CONCLUSION

The obvious feature of ETS education system is that the use of computer information technology platform replaces the traditional blackboard mode. First, the advanced computer system determines the scientific of paperless teaching system; and application of device identification code increase the credibility of ETS system. Computer can process and save data, comparing to traditional education model, so the test scores will get more accurate data. Moreover, the environmental paperless system has changed the traditional teaching model, such as environmental paperless record, examination and assignment be recorded online. Which save a lot of resources, but also improve the teaching efficiency. With the information age coming, computer and network applications become more and more common, ETS paperless teaching system is bound to spread widely, it will play a huge role in the conservation of resources and improve efficiency and teaching quality.

ACKNOWLEDGMENT

The authors would like to sincerely thank all personnel associated with this project for their past and continued support.

REFERENCES

- [1] Nikos Bozionelos. Socio-economic background and computer use: the role of computer anxiety and computer experience in their relationship, *International Journal of Human-Computer Studies*, Vol. 61, No. 5, November 2004, pp:725-746
- [2] Ramón Rubio García, Ramón Gallego Santos, Javier Suárez Quirós, et al. Present state of CAD teaching in Spanish Universities. *Computers & Education*, Vol. 44, No.3, April 2005, pp: 201-215
- [3] Yao Huantao. The exploration for paperless office on campus. *Read and write journal*, Vol.4, No.6, 2007, pp:10. (in Chinese)
- [4] Taha Mzoughi, S. Davis Herring, John T. Foley et al. WebTOP: A 3D interactive system for teaching and learning. *Computers & Education*, Vol. 49, No 1, August 2007, pp: 110-129
- [5] Zhang Yangqing. The exploration for the implementation of the paperless office in colleges and universities. *Computer Knowledge and Technology*, Vol. 5, No 35, 2009, pp:9986-9987. (in Chinese)
- [6] Wei Bin, Chen Weiwei. Design and development for paperless examination system. *Nanjing Xiaozhuang College*, Vol. 20, No 4, 2004, pp:75-78. (in Chinese)
- [7] Liu Liying. "College English" paperless teaching and examination nature. *Journal of zhenjiang tall*, 2006, Vol. 19, No 3, pp:118-120. (in Chinese)
- [8] Luo Dongjun. Research and practice based on Email job manage-

- ment. *Chengde Petroleum College newspaper*, 2007, Vol. 9, No 1, pp:39-44. (in Chinese)
- [9] Tian Geming. The implementation measures for paperless examination system to prevent cheating. *Sanming College*, 2007, Vol. 24, No 4, pp: 456-459.
- [10] Siu Cheung Kong. Using a web-enabled video system to support student-teachers' self-reflection in teaching practice *Computers & Education*, December 2010, Vol. 55, No 4, pp:1772-1782.
- [11] Qi Juhui. The discussion of some key technologies of online examination system. *Software Guide*, Vol. 10, No 9, 2011, pp:169-170. (in Chinese)
- [12] XuFeng. The experience of the development and implementation of network test system. *Chinese scientific and technological innovation Tribune*, 2008:210.
- [13] M.M El-Khouly, B.H Far, Z Koono. Expert tutoring system for teaching computer programming languages *Expert Systems with Applications*, Vol. 18, No 1, January 2000, pp: 27-32
- [14] Azizah Jaafar, Shahrir Abdullah, Azian Abd. Latif, et al. Reliability of Student Feedback on the Course Teaching Evaluation System (CTES) and System Usability *Procedia - Social and Behavioral Sciences*, Vol. 18, 2011, pp:24-32
- [15] Maryam Sedaghat, Soudeh Rahmani. A review of approaches to teaching thinking: appropriate approach for Iran education system. *Procedia - Social and Behavioral Sciences*, Vol. 30, 2011, pp: 1037-1042

Development and Application of Oil Sand

Xiaoming Zhang, Yi Pan
Liaoning Shihua University, Fushun, China

Abstract— the oil sands is one kind of sand-like unconventional ore resources depositing of crude oil, it is made up of asphalt, minerals, clay and water in a way of conjuncting with each other. Oil and natural gas are important energy and chemical raw materials, its resources are gradually reduced. With the rapid development of the global economy, the conventional oil resources can't meet the rapid growth of oil demand, people began turning to unconventional oil resources, one of which is the oil sands. Oil sands is unconventional oil resources, if its proven reserves are converted into oil, it will be much larger than the world's proven oil reserves. Canadian oil sands reserves stand ahead in the world, followed by the former Soviet Union, Venezuela, the United States and China. However, due to its special properties, different mining and processing technology, and higher mining costs compared with conventional oil, the research of oil sands makes slow progress. At present, due to the rising of world oil price, oil sands mining technology have attracted more and more attention, and have developed a lot.

Index Terms — oil sands, mining, alternative energy, application, energy resource, unconventional energy.

1 INTRODUCTION

The oil sands is one kind of sand-like unconventional ore resources depositing of crude oil, it is made up of asphalt, minerals, clay and water in a way of conjuncting with each other. Asphalt is its main ingredient, whose content can be accounted for 1% to 20%. Oil sands formed in Cretaceous. There are more than 70 countries in the world that reserve oil sands, but more than 90% of the proven oil sands reserves concentrate in Canada at present; if all these oil sands resource were exploited, according to the current level of the world energy needs, they can be used 100 years by the world. Oil sands asphalt is usually the organic mixture of hydrocarbon and non-hydrocarbon, it is a viscous semi-solid substance, containing about 80% of carbon, certain hydrogen, and a small amounts of nitrogen, sulfur, oxygen, and trace metals. Oil sands is one of the main sources of man-made oil. Through mining, extraction, separation, modification, we can get synthetic crude oil from it, and the separated clean sand can be used as construction materials or used for backfill process.

To modify the running headings, select View | Header and Footer. Click inside the text box to type the name of the journal the article is being submitted to and the manuscript identification number. Click the forward arrow in the pop-up tool bar to modify the header or footer on subsequent pages.

2 EXPLOITATION TECHNIC

Oil sands differ greatly from conventional oil and natural gas, in the original state, it is difficult to use ordinary methods to mine. In order to exploit it, special treatment is needed. According to the differences of the depth of oil sand reserves, the mining methods of oil sands can be divided into in situ mining and open-pit mining. The combining of thermal recovery and chemical flooding method can improve the production rate and oil recovery rate, it is a promising method. Hot alkaline extraction method is the most widely used method of open-pit mining, but it will cause serious environmental pollution. When combining with surfactant, this method can improve the recovery rate of asphalt, lower production costs, reduce environmental pollution.

2.1 Open-pit Mining

The process of open-pit mining is: first, remove the oil sands on the first cover, and then mine by the open-pit mining techniques. The exploited oil sands were then shipped to the factories to separate asphalt, sand and other non-hydrocarbon substances. And finally, convert the separated asphalt into synthetic crude oil. Surface mining method can take out more than 90% of the oil. The hydrocarbons, sulfur, metals and rocks in the oil sands have been utilized to the fullest. There are only 10% of the oil sands around the world that can use open-pit mining technology to mine. Open-pit mining has a better asphalt recycle rate than other methods, and it is technically more mature. There have formed large-scale industrial mining in Canada and Venezuela. The Syncrude company in Canada is the world's largest manufacturer that produces oil from oil sands, who engaged in the open-pit mining activities in the Athabasca. Its open-pit mining technology of oil sands leads the world.

2.2 In-situ mining

The site separation method that is used to exploit deep underground oil sands has also been very mature. This approach is to heat and dilute the asphalt to be flowable in the underground, and then pump the flowable asphalt to the ground. Nowadays, widely used methods are as following:

- 1) Cyclic steam stimulation (CSS): cyclic steam stimulation, also known as cyclic steam injection, steam soak, or steam production and so on, is injecting high-pressure steam into the oil sands formations, shut-in for a few weeks, use heat to soften the asphalt, use the water vapor to dilute the bitumen and make the asphalt and sand separate, and then open the wells to pump the bitumen that can flow to the surface[1]. In addition, through the application of chemicals, the viscosity of heavy crude oil will reduce, this can improve heavy oil's mobility in the reservoir and on the ground, improve the state of steam stimulation, enhance oil recovery [2]. Its main advantage is the immediate producing of oil after the project be implemented, while its main limitation is that it can exploit

20% of the original underground reserves[3]. Steam-assisted gravity drainage method (SAGD): Drill two wells in parallel to the oil sands formations, one up, one down, then continuously injecting steam to the upper well, when the steam heat the oil formation, oil Sands will be softened and flow by gravity to the below well, then be pumped to the ground. At present, this production technology can harvest 25% to 75% of the underground storage of asphalt, which is far greater than conventional light oil recovery. UTF Consortium's application in ForMt cMurray area is SAGD technology. The future development of oil sands is expected to mainly depend on the combination of CSS technology and SAGD technology, which will surely become the 21st century's major commercial mining method.

- 2) gy developed in Canada in recent years is a new mining technology. To the underground heavy oil t steam-assisted gravity drainage method (SAGD) [4]: Drill two wells in parallel to the oil sands formations, one up, one down, then continuously injecting steam to the upper well, when the steam heat the oil formation, oil Sands will be softened and flow by gravity to the below well, then be pumped to the ground. At present, this production technology can harvest 25% to 75% of the underground storage of asphalt, which is far greater than conventional light oil recovery. UTF Consortium's application in ForMt cMurray area is SAGD technology. The future development of oil sands is expected to mainly depend on the combination of CSS technology and SAGD technology, which will surely become the 21st century's major commercial mining method.
- 3) Vapor extraction technology of the underground horizontal wells (VAPEX): This method is a further development of the steam assisted gravity drainage method. In VAPEX, hydrocarbon gas is injected rather than steam. The main advantage of the process is: compared with SAGD, VAPEX's necessary equipments are cheap, simple to handle, and the gas dissolves selectively. It only dissolves in the oil, not in the water. So this method can apply to a wide range.
- 4) Cold heavy oil production technology (CHOP): the cold heavy oil production technology has certain mobility, cold heavy oil production wells can greatly improve the wells' ability of producing conventional oil. In the Inner Mongolia oil fields of China, the cold heavy oil production technology is relatively successful.
- 5) Aquathermolysis mining technology: In this mining technology, under the conditions of injecting steam, by the help of chemical reactions that occurs between heavy oil and steam, the viscosity of heavy oil will reduce, thus achieving the purpose of lower viscosity underground mining.
- 6) Recovery heavy oil by in-stu reducing viscosity techniques: In simple thermal effects, thermal cracking and visbreaking will happen to the heavy oil and its quality will change. Therefore some people proposed that in the heat treatment, injecting hydrogen and other substances that can produce hydrogen element to the underground. In addition to integrating into the crude oil to make the oil expand, the hydrogen can enter the tiny pores; when the underlying pressure re-

duce, the expansion of the compressed gas and crude oil can produce more oil, and heat the crude oil at the same treatment. Due to expansion, the viscosity of crude oil will reduce.

- 7) In Situ combustion: In Situ combustion uses electricity and chemical methods such to burn oil reservoir to increase its temperature to burn point. In the meantime, inject air to allow the crude oil to burn continuously[5]. The disadvantages include that the implementation of in situ combustion process is difficult. Controlling the underground combustion is also not easy. Also, the cost of injecting a mass of high-pressure air into the reservoir is high. Usually, this method uses when steam flooding method is not applicable, especially in the thin layers or deep layers whose crude oil is of medium density.

3 OIL SEPARATION TECHNOLOGY

Currently, there are three main kinds of oil sands separation technology at abroad-hot water washing method, solvent extraction method, pyrolysis distillation method. Different oil sands structures may use different separation methods. Generally, moist oil sands are for washing method, oily sand for organic solvent extraction or pyrolysis distillation separation. Thermo-chemical washing method and distillation ATP will be the main ground separation methods of oil sand in the future, which are still dominated by chemical washing. But with the rising of the oil prices, ATP carbonization technology will have a period of vigorous development [5].

Domestic oil sands separation technology is still stay in the laboratory study stage. To the oil sands of Inner Mongolia, Cao Zubin, Professor of Liaoning Shihua University, proposed water-washing separation technology, and had developed three series of washing separation reagents. The results prove to be very ideal. The dry distillation simulation test that was conducted on Xinjiang oil sands by Langfang Branch of China Petroleum Exploration and Development Research Institute had also achieved desired gains, which proved that the dry distillation process of Xinjiang oil sand is technically feasible.

3.1 Hot water washing method of Oil sands

Currently, the Canadian oil sands surface separation mainly uses hot water / surfactant. Hot alkali and surfactants role in changing the surface wettability of the sand, making the surface of sand more hydrophilic, achieving the separation of asphalt absorbing on sand. The separated asphalt oil will then float to the lye, while the quartz sand will take the lower part, thus achieving the objective of separation.

3.2 Organic solvent extraction of oil sand

Pyrolysis s Organic solvent extraction of oil sands mainly uses the theory that similarities can be solvable easily in each other to achieve the recycle of oil sands bitumen. Using organic solvent to extract the gel bitumen in the sand, and then distill them to achieve separation. This method makes use of organic solvents to contact with the oil sands, separating the dissolved asphalt from

sand. And the extraction agent can be recycled. Solvent extraction method applies to theoretical and laboratory research stage. This method is less demanding on the quality of oil sands, it can be wet oil sands with high oil content, it can also be sands with little oil, or just dry oil sands. Compared with the hot water washing method, its applicability is broader, and washes oil more efficiently.

Disadvantage of this approach and also its largest problem, is serious environmental pollution. Therefore, this method is rarely used in industrial production.

3.3 Pyrolysis Separation Method of Oil sands

About the pyrolysis separation technology—the main idea of Canadian Alberta's Oil Sand Technology Administration is transforming the heavy oil sands components to be light components. Aostar Taciuk Process referred to as the ATP process, the principle is the use of high temperature of above 250e to pyrolysis. After the heat treatment, greatly improving the quality of asphalt improves greatly. Molecular mass becomes smaller, the quantity of gum also reduces. During the heat treatment, the most important change is the emerging of light oil.

In the 1990s, Canada began to apply ATP technology to oil sands mining. AOSTA and UMATAC cooperated to make efforts in the feasibility studies of the oil sands pyrolysis. In the past 17 years, the technology went through continuous development and improvements. The ATP pilot plant in southeast Calgary has processed more than 1.7 @ 104t oil sands, showing the ATP method is a technically effective way for separation and primary modification.

4 PROCESSING TECHNOLOGY OF OIL SANDS

The key to the processing routes and process selection of oil sands asphalt is the technical and economic feasibility. To the characteristics of oil sands asphalt, Xing Dingfeng and his colleagues, probe and conclude that the processing of oil sands asphalt mainly has decarbonization and hydrogenation two routes.

4.1 Hydrogenation of oil sands bitumen

On the hydrogenation routes, according to different reactors, the heavy oil hydrogenation process can be divided into fixed-bed, boiling bed, suspending bed and moving bed 4 species. Nowadays the application of fixed bed and boiling bed technology is relatively mature. But the further processing of oil sands bitumen uses residual oil boiling bed hydrogenation cracking technology.

Currently, the oil sands bitumen hydro-processing technology of Canadian mainly employs boiling bed technology. One of the typical boiling bed hydrogenation cracking technology is the LC-Fining process. Except the processing of conventional sour crude oil. This technology mainly uses to process heavy residual oil. In this processing scheme, considering the maturity of the technology, the scale of the completed device and the serviceability of its processing raw materials, LC-Fining technique is recommended.

4.2 Decarbonization process of oil sands asphalt

Decarbonization routes mainly include solvent de-asphalting, catalytic cracking and coking process, but it seems that only the

coking technology is feasible for the processing of oil sands asphalt at present. Residue oil coking technology is the primary means for the deep processing of residue oil and for improving the recycle rate of light oil. Currently, the technology is already very mature, including delayed coking, fluid coking and flexible coking. Coking process can process a variety of low-quality raw material, it is unlimited to materials' properties.

5.CONCLUSION

Current oil sands development and utilization is just limited to a few countries, while Canada is the largest oil sands producer among them. Although the current cost of oil sands mining and refining is high, with the development of technology and global oil prices will remain at a high level of expectations, the economic value of extracting oil from oil sand will increase.

The depletion of conventional oil resources, makes the huge gap of future energy largely depend on oil sands to make up. The world began to realize the importance of oil sands resources, and start pouring a lot of manpower and material to develop this unconventional energy sources. And, the advances in mining technology will allow the extraction of some oil sands and ores that do not have the economic value before also economically possible.

In the conditions that domestic oil resources is tense, be familiar with the state of our oil sands resources and preparing for the exploitation of oil sands resources will make sense to China's energy strategy.

REFERENCES

- [1] Zhang Chunhe, major in the future development direction of the world oil industry's transition [A]. The Chinese mining industry, 2005,14 (7): 14 – 15.
- [2] Gao Zhinong, Xu Donghua, the physical and chemical properties of Qinghai oil sands [J] Instrumental Analysis newspaper, 2002, 21 (6): 1-4.
- [3] J.O'M. Bockris. Will lack of energy lead to the demise of high-technology countries in this century?[J]. International Journal of Hydrogen Energy, 2007:153-158.
- [4] Chunhe zhang. Transitional trend of world oil industry in coming years[J]. China Mining Magazine, 2005, 14(7): 13-15.
- [5] <http://baike.baidu.com/view/2642417.htm>
- [6] B.D. Sparksa, L.S. Kotlyara, J.B. O'Carrolla, K.H. Chungb. Athabasca oil sands: effect of organic coated solids on bitumen recovery and quality[J]. Journal of Petroleum Science and Engineering, 2003, 417-430.
- [7] Xing Dingfeng; et al—Study on Processing Schemes of Canada Oil-sand Bitumen[J] Petroleum Planning & Engineering, 2007, 01.
- [8] CUI Miao-miao, LI Wen-shen; LI Xiao-ou, FENG Rui-jiang, Development and Utilization of Oil Sands [J]. Chemical Industry and Engineering, 2009, 01.
- [9] ZHANG Yi-ge, CAO Zu-bin¹, YANG Fan, XU Xiu-qiang¹, WORLD DEVELOPMENT STATUS OF OIL-SAND SEPARATION TECHNOLOGIES[J]. Natural Gas Industry, 2008, 12.
- [10] Sun Weiwei, Zhao Dezhi, Sun Mingzhu, Oil Sand and Separation Method [J]. China Chemicals, 2007, 05.
- [11] ZHENG De-wen, FANG Chao-he, LI Jian, GE Zhi-xin, WANG Yi-feng, SUMMARY ON OIL SAND EXPLOITATION TECHNOLOGIES AND METHODS [J]. Journal of Southwest Petroleum University(Science & Technology Edition), 2008, 06.
- [12] Li Shuyuan; Wang Jianqiu; Qian Jialin, Research, Development and Utilization of Oil Sand Resources in the World [J]. Sino-Global Energy, 2011, 05.
- [13] SHAN Xuan-long¹, CHE Chang-bo; LI Jian; FAN Chao-ying, ZANG Chun-yan, WANG Qing-bin, Present status of oil sand resources at home and abroad [J]. Global Geology, 2007, 04.

- [14] Li Shuyuan, Wang Jianqiu, Tan Huaping and Wu Zhaoliang. Study of extraction and pyrolysis of Chinese oil sands [J]. Fuel, 1995, 74(8): 1191-1193.
- [15] Wang Kaiming. Exploration and Development Prospects of Unconventional Hydrocarbon Resources in China[J] Petroleum & Petrochemical Today. 2009, 04.
- [16] FENG Xue-wei, CHENG Chen, CHEN Da-yong. New development of oil shale in-situ technology [J]. China Mining Magazine, 2011, 06.

Power Distribution of Wind Diesel Generator in Isolated Network

Prashant Bawaney, B. Sridhar

Abstract— Recent research and development of alternative energy sources have shown excellent potential to complement the contribution to conventional power generation systems. In order to meet sustained load demands during varying natural conditions, different energy sources and converters need to be integrated with each other for extended usage of alternative energy. This paper focuses on the combination of Wind Turbine and Diesel Generator systems for sustained power generation. As the wind turbine output power varies with the wind speed: a DG system can be integrated with the wind turbine to ensure that the system performs under all conditions. This paper presents Dynamic Modelling and Simulation of a Wind/DG Hybrid Power Generation System with power flow controllers.

Index Terms: point of common coupling (PCC), squirrel-cage induction generator (SCIG), Wind-Diesel system, Isolated network, Variable-Speed wind Turbines, DC Exciter, Discrete Frequency regulator.

1. INTRODUCTION

Wind energy is gaining increasing importance throughout the world. This fast development of wind energy technology and of the market has large implications for a number of people and institutions: for instance, for scientists who research and teach future wind power and electrical engineers at universities; for professionals at electric utilities who really need to understand the complexity of the positive and negative effects that wind energy can have on the power system; for wind turbine manufacturers; and for developers of wind energy projects, who also need that understanding in order to be able to develop feasible, modern and cost-effective wind energy projects. See Table 1 for an overview of important historical wind turbines.

TABLE 1
 Over View of Important Wind Turbine

Turbine and country	Diameter (m)	Swept Area (m ²)	Power (kW)	Specific Power (kW/m ²)	Number of blades	Tower height (m)	Date in service
Poul LaCour, Denmark	23	408	18	0.04	4	—	1891
Smith-Putnam, USA	53	2231	1250	0.56	2	34	1941
F. L. Smidth, Denmark	17	237	50	0.21	3	24	1941
F. L. Smidth, Denmark	24	456	70	0.15	3	24	1942
Gedsder, Denmark	24	452	200	0.44	3	25	1957
Hütter, Germany	34	908	100	0.11	2	22	1958

- Prashant Bawaney, Assistant Professor, Electrical Engineering Department, Christian College of Engineering & Technology, Bhilai, Chhattisgarh, India, PH-098274-62782, E-mail: pbawaney@gmail.com
- B. Sridhar, Assistant Professor, Electrical Engineering Department, Christian College of Engineering & Technology, Bhilai, Chhattisgarh, India, PH-097527-27860, E-mail: burlasridhar@gmail.com

2. CONCEPTS OF WIND POWER GENERATION

The main components of a wind turbine system, including the turbine rotor, gearbox, generator, transformer, and possible power electronics, are illustrated in Fig. 2.1. The wind turbines to be connected at remote sites with high average wind speed. Such sites are often situated far from a strong grid [1]. The turbine rotor converts the fluctuating wind energy into mechanical energy, which is converted into electrical power through the generator, and then transferred into the grid through a transformer and transmission lines. Wind turbines capture the power from the wind by means of aerodynamically designed blades and convert it to rotating mechanical power. The number of blades is normally three and the rotational speed decreases as the radius of the blade increases. For MW range wind turbines the rotational speed will be 10–15 rpm. The weight efficient way to convert the low-speed, high-torque power to electrical power is to use a gearbox and a generator with standard speed. The gearbox adapts the low speed of the turbine rotor to the high speed of the generator. The gearbox may be not necessary for multi-pole generator systems.

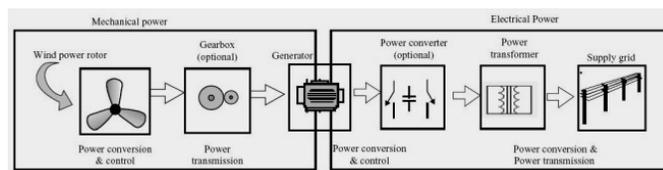


Fig 2.1 General Wind Power Generation

The generator converts the mechanical power into electrical energy, which is fed into a grid through possibly a power electronic converter, and a transformer with circuit

breakers and electricity meters. The generator is coupled to the grid through a transformer and/or a power electronic converter, because the characteristics of the generator output do not match the characteristics of the grid with respect to frequency and voltage [2]. The connection of wind turbines to the grid is possible at low voltage, medium voltage, high voltage, and even at the extra high voltage system since the transmittable power of an electricity system usually increases with increasing the voltage level. While most of the turbines are now a days connected to the medium voltage system, large offshore wind farms are connected to the high and extra high voltage level. The amount of power produced by a turbine can be expressed as $P = 0.5\rho A C_P V^3$ [3]. A high rated wind speed will give a large peak-power output for a particular swept area, and may give a greater overall energy production [4].

At the point of common coupling (PCC) between the single wind turbines or the wind farm and the grid, there is a circuit breaker for the disconnection of the whole wind farm or of the wind turbines. Also the electricity meters are installed usually with their own voltage and current transformers. The electrical protective system of a wind turbine system needs to protect the wind turbine and as well as secure the safe operation of the network under all circumstances.

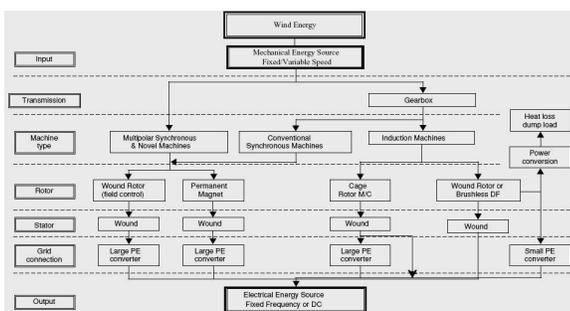


Fig 2.2 Roadmap for wind energy conversion

The possible technical solutions of the electrical system are many and Fig. 2.2 shows a technological roadmap starting with wind energy/power and converting the mechanical power into electrical power. Torque peaks in the gearbox and shafts are reduced, the wind turbine can operate with maximum aerodynamic efficiency and power fluctuations can be absorbed as inertial energy in the blades [5]. It involves solutions with and without gearbox as well as solutions with or without power electronic conversion.

3. GENERATOR SYSTEMS FOR WIND TURBINES

Both induction and synchronous generators can be used for wind turbine systems. Induction generators can be used in a fixed-speed system or a variable-speed system, while synchronous generators are normally used in power electronic interfaced variable-speed systems. Mainly, three types of induction generators are used in wind power conversion systems: cage rotor, wound rotor with slip control by changing rotor resistance, and doubly fed induction generators. The wound rotor generator with rotor-resistance-slip control is normally directly connected to an ac system, but the slip control provides the ability of changing the operation speed in a certain range. The doubly fed induction generators provide a wide range of speed variation depending on the size of power electronic converter systems. In this paper we discuss the systems without power electronics except the thyristor soft starter, and the variable-speed wind turbine systems, including those with partially rated power electronics and the full-scale power electronic interfaced wind turbine systems.

3.1 Fixed-Speed wind Turbines

In fixed-speed wind turbines, the generator is directly connected to the mains supply grid. The frequency of the grid determines the rotational speed of the generator and thus of the rotor. The generator speed depends on the number of pole pairs and the frequency of the grid. The Danish Concept, of directly connecting a wind turbine to the grid, is widely used for power ratings up to 2.3 MW. The scheme consists of a squirrel-cage induction generator (SCIG), connected via a transformer to the grid. The wind turbine systems using cage rotor induction generators almost operate at a fixed speed (variation of 1–2%). The power can be limited aerodynamically by stall control, active stall control, or by pitch control.

3.2 Variable-Speed wind Turbines.

In variable-speed systems the generator is normally connected to the grid by a power electronic system. All variable speed generators need converters to control the voltage and frequency of power supplied to the grid [6]. Only the rotor of the generator is connected through a power electronic system. This means the nominal power of the converter system can be less than the nominal power of the wind turbine. By controlling the active power of the converter, it is possible to vary the rotational speed of

the generator and thus of the rotor of the wind turbines. The active power flow is represented by the wind turbine power curve which is the relation between the active power produced and the wind speed [7]. For a variable speed wind turbine with a doubly fed induction machine, it is possible to control the load torque at the generator directly, so that the speed of the turbine rotor can be varied within certain limits [8].

4. MODEL DESCRIPTION

A model of the, Wind-Diesel system is presented in this Modeling and Simulation in this paper. The optimal wind penetration (installed wind capacity/peak electrical demand) for this system depends on the site delivery cost of fuel and available wind resource. The wind-diesel system presented in this demo uses a 300 kVA synchronous machine, a wind turbine driving a 275 kVA induction generator, a 75 kW customer load and a variable secondary load (0 to 446.25 kW). Simulation model shown in fig. 4.1.

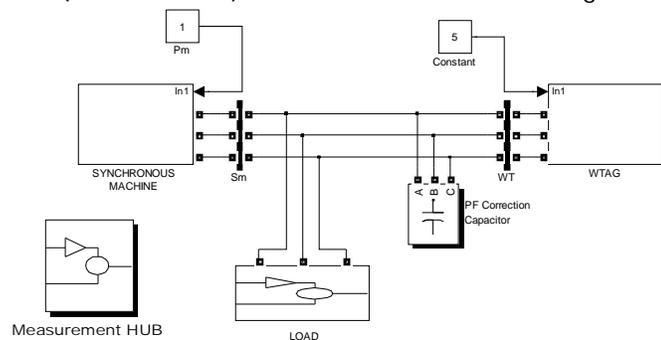


Fig 4.1 Simulation model

4.1 Diesel Generator

The mechanical power input to the synchronous machine which acts as a synchronous generator when the WTAG works below cut in speed is fed from a constant block of value 1 PU as shown in fig 4.2. There is also need of proper excitation system fig 4.3 shows a predefined model of simulink for excitation system. The Excitation System block is a Simulink system implementing a DC exciter. The basic elements that form the Excitation System block are the voltage regulator and the exciter. The output of the block is the field voltage V_f in pu, to be applied to the V_f input of a synchronous machine block.

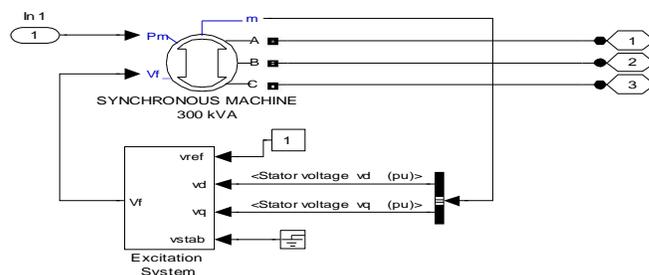


Fig 4.2 Simulink model for Diesel Generator

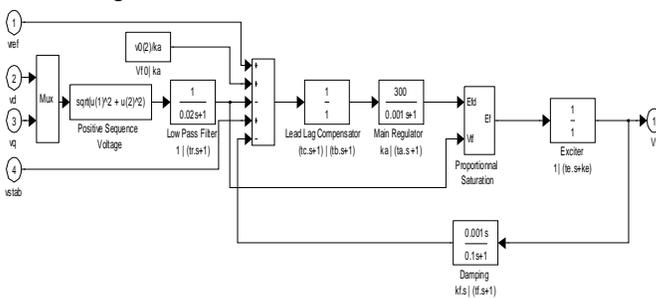


Fig 4.3 Excitation System

4.2 Wind Turbine Asynchronous Generator

The Wind Turbine Asynchronous Generator model is much of complicated part of whole simulation model. It consists of Asynchronous Generator & Wind Turbine. Wind Turbine Asynchronous Generator shown in fig. 4.4. The Wind Turbine block uses a 2-D lookup table to compute the turbine torque output (T_m) as a function of wind speed (w_{Wind}) and turbine speed (w_{Turb}). The lookup table graph between wind speed and turbine speed gives mechanical power. The mechanical torque is obtained by dividing mechanical power with wind speed. This mechanical torque is the input for asynchronous generator.

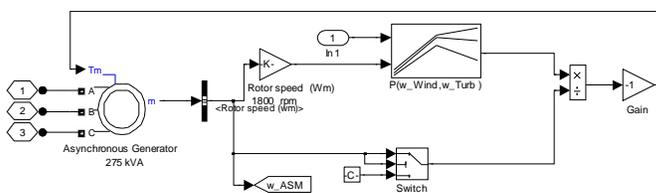


Fig 4.4 Wind Turbine Asynchronous Generator

4.3 Load Centre

Load center consist of three major part, Main load (50 kw), Secondary Load (0 -446.25 kw) & Auxiliary Load (25 Kw) . All loads which are constant running are assumed to be in Main Load (Base Load) & Auxiliary load

is a fluctuating load with 25 kW i.e. its that type of load which is not necessary to be in connected condition all the load for all time. While Secondary load is variable load which consumes extra generated power from wind turbine asynchronous generator and synchronous machine & dumps the fluctuated frequency.

In this load center, we used two circuit breakers for connecting and disconnecting auxiliary load and secondary load. When we want to control the frequency we close the circuit breaker of secondary load and vice versa. Fig. 4.5 shows Simulink model of Load Centre.

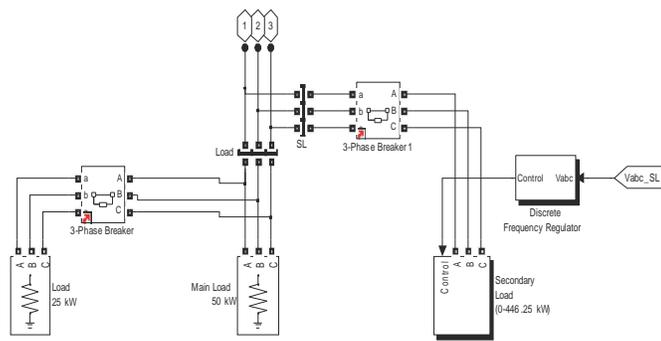


Fig 4.5 Details of Load Centre

4.4 Measuring Systems

Every Simulation requires results and results can pull out from measurement system. Hence explanation and details of it are shown in fig 4.6. At low wind speeds both the induction generator and the diesel-driven synchronous generator are required to feed the load. When the wind power exceeds the load demand, it is possible to shut down the diesel generator. A secondary load bank is used to regulate the system frequency by absorbing the wind power exceeding consumer demand.

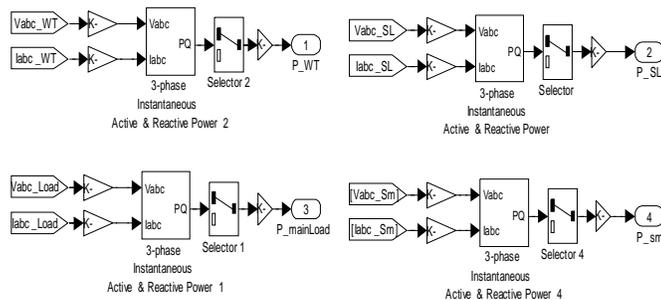


Fig 4.6 Measurement system

5 Results

The wind speed (0- 9 m/s) is such that the wind turbine produces enough power to supply the load. The synchronous machine operates with its mechanical power input (Pm) set at zero. The dynamic performance of the frequency regulation system when an additional 25 kW customer load is switched on. As the asynchronous machine operates in generator mode, its speed is slightly above the synchronous speed. Two different test cases are done for understanding the effect of all parameters on the

isolated system. The case wise results are shown in this report as follows.

CASE-1 The mechanical power input (Pm) of synchronous machine is zero pu with the frequency control.

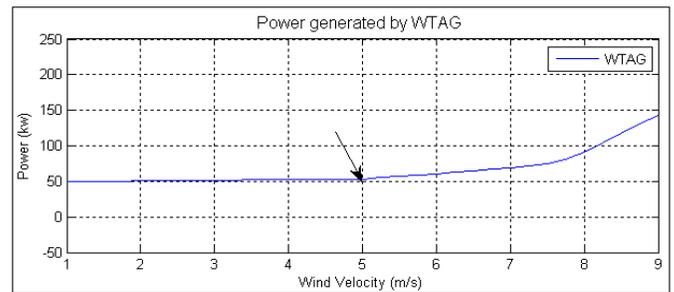


Fig 5.1 Graph between wind velocity and power of Wind Turbine Asynchronous Generator of case 1

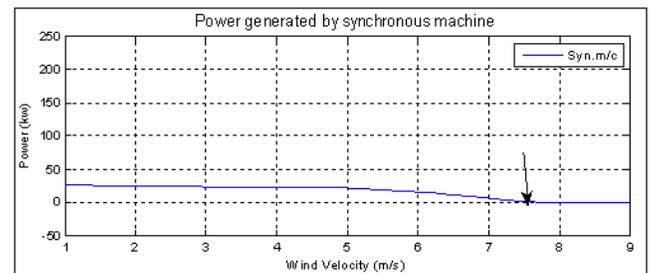


Fig 5.2 Graph between wind velocity and Power of Synchronous machine of case-1

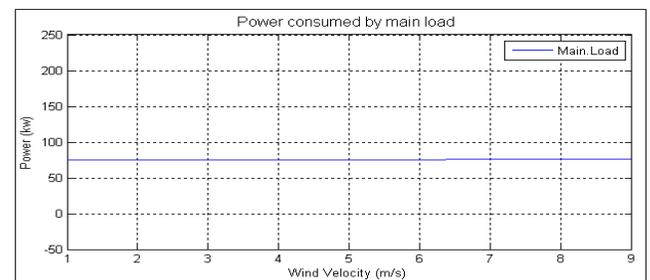


Fig 5.3 Graph between wind velocity and Power of Main Load of case-1

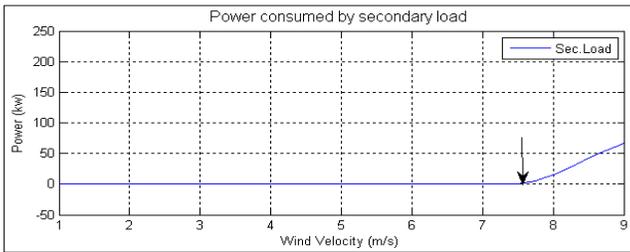


Fig 5.4 Graph between wind velocity and Power of Secondary Load of case-1

Fig 5.7 Graph between wind velocity and power of Main Load of case -2

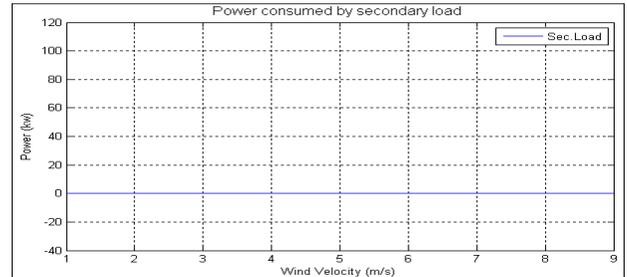


Fig 5.8 Graph between wind velocity and power of Secondary Load of case -2

CASE-2 The mechanical power input (P_m) of synchronous machine is zero pu without the frequency control.

6. Conclusions

As a part of paper, two major case studies are handled in this paper. All case studies are done with the intent to understand the behavior of whole system in different conditions. Results and conclusion drawn from simulations leads us towards need of studying and designing more efficient controllers in generation side as well as load center side. Though load center controller is simulated in this paper, it is required to investigate more robust controller.

Acknowledgment

I would like to express to my gratitude sincere thanks to all people who have taken a great deal of interest in my manuscript and helped me with valuable suggestion that helped me a lot. First and foremost I thanks to **Prof. S. C. Tiwari**, HOD of the Electrical Engineering department for keeping the faith in me and continually keeping me encouraged by his valuable suggestion. I acknowledge with gratitude the benediction of CCET, Bhilai, **Dr. Fr. Jose. K. Varghese**, Executive Vice Chairman CCET, Bhilai and **Dr. R.N. Dash**, Director, CCET, Bhilai to extend all facilities and Co-operation in completion of this manuscript.

References

- [1] Nayeem Rahmat Ullah, " Small Scale Integration of Variable Speed Wind Turbines into the Local Grid and Its Voltage Stability Aspects", Future power system 2005 international conference, pp 8, Nov.2005.
- [2] J. G. Sloopweg, H. Polinder and W.L. Kling, " Dynamic Modelling of a Wind Turbine With Doubly Fed Induction Generator", Power engineering society

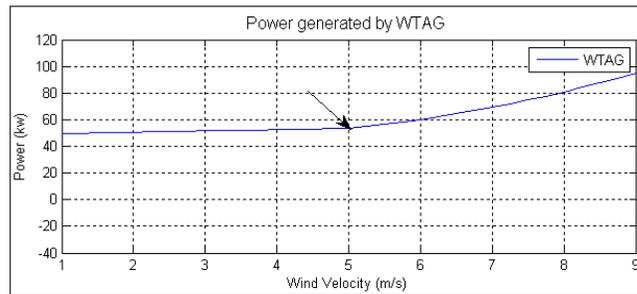


Fig 5.5 Graph between wind velocity and power of Wind Turbine Asynchronous Generator of case -2

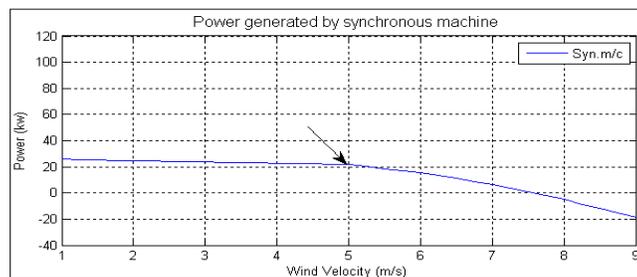
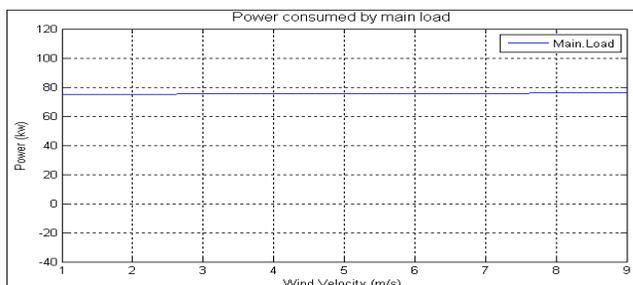


Fig 5.6 Graph between wind velocity and power of Synchronous machine of case -2



- summer meeting, Volume-1, pp 644-649,2001.
- [3] Roohollah Fadaeinedjad, Gerry Moschopoulos and Mehrdad Moallem, "Flicker Contribution of a wind Turbine in a Stand-Alone Wind Diesel System", Electrical and computer engineering, CCECE 2008, Canadian conference, pp 233-238, May 2008.
- [4] D.G. Infield, Prof. N. H. Lipman & P. J. Musgrove, "Review of wind/diesel Strategies", Physical science, measurement and instrumentation, management and education, reviews, IEE proceedings, Volume 130, Pt. A, No. 9, pp 613 – 619, Dec. 1983.
- [5] Roberto Cbrdenas Rub Cn Peiia Jon Clare & Greg Asher, "Power Smoothing in a Variable speed Wind-Diesel System " , Power electrncs specialist conference , PESC 03 , IEEE 34th annual, Volume 2, pp 754 – 759, June 2003.
- [6] P. A. Stott and M. A. Mueller, " Modelling Fully Variable Speed Hybrid Wind Diesel Systems", Universities power engineering conference,UPEC 06 Proceedings of the 41st international Volume 1, page 212 – 216, Sept. 2006.
- [7] Wellington Lemos, Alexandre Pereira, Renato Barros and Everaldo Feitosa,"Problems of planning Hybrid Wind–Diesel Power Systems", IEE /PES Transmission & Distribution Conference & Exposition, pp 617–622, Nov. 2004.
- [8] Tao Sun , Zhe Chen &FredeBlaabjerg, "Flicker Study on Variable Speed Wind Turbines With Doubly Fed Induction Generators",IEEE Transactions On Energy Conversion, Volume 20, NO. 4, pp 896 – 905, Dec. 2005.
- [9] Ongning Chi, Yanhua Liu,Weisheng Wang & Huizhu Dai , " Voltage Stability Analysis of Wind Farm Integration into Transmission Network", International Conference on Power System Technology, pp 1-7, Oct. 2006.
- [10] Atul S. Kini & Udaykumar R. Y., "Modelling And Performance Analysis of A Wind/Diesel Hybrid Power System ", Industry applications conference, 41st IA Annual meeting conference Volume 3, pp 1215 – 1221, Oct. 2006.
- [11] L. Shridhar Bhim Singh, C. S. Jha and B. P. Singh, "Analysis Of Self Excited Induction Generator Feeding Induction Motor ", IEEE Transactions on Energy Conversion, Vol. 9, No.2, pp 390 – 396, June 1994.
- [12] Eduardo Sukez and Gustavo Bortolotto "Voltage-Frequency Control Of A Self Excited Induction Generator " , IEEE Transactions on Energy Conversion, Vol. 14, No. 3, pp 394 – 401, Sep. 1999.

FitzGerald's Approach to Translation

Habibollah Mashhady

Mahbube Noura

Abstract

The present paper attempts to explore FitzGerald's overall approach to translation by examining his translated works and particularly by focusing on his translation of Khayyam's Rubaiyat. Khayyam Rubaiyat is selected as the text to gather data and it is compared to its English translation by FitzGerald in order to identify the strategies he used in translating it into English. In order to do so, first this paper traces translation theories common in Victorian era and the extent FitzGerald adheres to them. Second it discusses translation of poetry and potential obstacles inherent in it according to form, meaning, culture-specific terms and the role of interpretation in poetry translation. Then it focuses on Khayyam Rubaiyat translation and identifies strategies used in its translation which ranged from addition, selection, omission, domestication to foreignization, all of which transformed Khayyam into a Epicurean western-poet philosopher. It further concludes that FitzGerald has employed the same strategies and theory in all of his translated works. In addition to Persian language and literature, It is useful to examine his approach to translation in his non-Persian translated works like Greek literature and make a comparison among Persian and non-Persian translations.

KeyWords: VictorianEra, FitzGerald, Khayyam, Translation, Translation Theory, Translation Strategies.

Introduction

Throughout history, written and spoken translations have played a crucial role in interhuman communication especially in providing access to important texts for scholarship and religious purposes; therefore, translation has a key role in transferring the cultural heritage of one nation to the other one and in the development of cultural and intellectual life. Although there is somehow consensus among the scholars about the importance of

translation in this respect, studying the history of translation from Cicero up to now shows the main fact that there has not been any agreement among scholars on how to translate (literal, faithful, sense-for sense). Because of this discrepancy among the scholars' viewpoints on translation methods, here it is important to notice that in some periods, one method of translation is more dominant although there are some exceptions. From the chronological point of view, the viewpoints on translation and the approaches exerted by translators in nineteenth century were different from the previous and next centuries. With a hindsight to the Victorian views on translation in nineteenth century, we try to explore the FitzGerald approach to translation in general and then examine his approach to translation in Khayyam Rubaiyat and find out strategies he used in translating it.

Methodology

In order to do this study, the book entitled *The Wine of Nishapour* by Shahrokh Golestan, the collection of Khayyam Rubaiyat, and its equivalent translation by Edward FitzGerald (1859) has been selected to explore FitzGerald's approach to translation and strategies he exerted in translating it. Since I am not competent in other languages like Greek and I know only English and Persian as my native language, I used only Persian translated works by FitzGerald. To do this study, I made one-to-one comparison between Persian quatrains of Khayyam and FitzGerald translation and in each case, the strategy used in translation has been identified. In addition to this type of work, I looked through the articles and books both Persian and English to use other ideas about FitzGerald's approach to the translation of texts from other languages, namely, Greek language.

Victorian Era and Translation

The Victorian Era of the United Kingdom was the period of Queen Victoria's rule from June 1837 to January 1907. This was a long period of prosperity for the British people, as profits gained from the overseas British Empire, as well as from industrial improvements at home. The era was preceded by the Georgian period and succeeded by the Edwardian period. The era is

often characterized as a long period of peace and economic, colonial and industrial consolidation. At that time, Iran was under the Empire of British which had a direct effect on the status British people devoted to the Persian literature. They regarded it as an inferior and poor literature which needs to be enriched by British scholars and we can see how this view impacts on the way they translate Persian literature at that time. The main principle of translation common in this era is the need to convey the remoteness of the original in time and place. It means that the original text is perceived as property, as an item of beauty to be added to a collection, with no concessions to the taste or expectations of contemporary life; so, it leads to archaic translation which satisfy the minority group; the translator focuses on the SL text; the TL reader is brought to the SL text. The nineteenth century brought new standard of accuracy and style in translation.

Regarding the style, the Victorian aim achieved through far-reaching metaphrase (literality) or pseudo-metaphrase was to constantly remind readers that they were reading a foreign classic. An exception was the outstanding translation of Khayyam Rubaiyat by FitzGerald in these periods which achieved its oriental flavor largely by using Persian names and discrete Biblical echoes and actually drew little of its material from the Persian original. In this period people were interested in theatre, arts, music, drama and opera as well as gambling, drinking and prostitution. In terms of technology and natural sciences the book entitled '*on the origins of species*' by Charles Darwin was written and published in this era. All of these events had a direct effect on the reception of the translation of Khayyam Rubaiyat and the strategies FitzGerald used in its translation. Before moving on FitzGerald overall approach to translation, it is appropriate to mention some factors important in poetry and hindrances in its translation.

Translation of poetry

Poetry presents special challenges to translators given the importance of a text's formal aspects in addition to its content. In his influential 1959 paper "on Linguistic Aspects of Translation", the Russian-born linguist and semiotician, Roman Jakobson went so far as to declare that

poetry by definition is untranslatable. In 1974 the American poet James Merrill wrote a poem "Lost in Translation" which in part explore this idea. So a good translation of a poem must convey as much as possible not only of its literal meaning but of its form and structure (meter, rhyme, alliteration, scheme, etc). Here we are going to discuss translation of poetry from three aspects: Form and meaning, culture-specific terms and interpretation of poetry.

1) **Form and meaning:** Translation of poetry is probably the subject in translation that triggers the strongest polemics. The problems originate from the multiplicity of meaning in a literary text also from the integration of form and meaning. Since the form and style of poetry is very imaginative and complex, it is very difficult, sometimes impossible to transfer all the linguistic features of a poem from one language to another. The form contains part of the meaning so that a loss in transferring the form leads to a loss in transferring the total meaning. So translator should be aware of all these features and integrity of form and meaning and should be competent to transfer or reproduce all these features in the target text.

2) **Culture – specific terms:** Cultural elements in translation of poetry have significant role. Thorough knowledge of a foreign language, its vocabulary, grammar is not sufficient to make one competent as a translator. One should be familiar with ones own culture and be aware of the source-language culture before attempting to build any bridge between them.

Nida(1964:55) classifies the cultural references in five groups:

1. Material, related to every day objects.
2. Ecological, related to differences in the places, weather...
3. Social related to social organization and its artistic manifestation in the arts or literature and history.
4. Religious
5. Linguistic, the tool which is needed to express the previous types of reference.

Based on Newmark, culture is the way of life and its manifestation that is peculiar to a community that uses a particular language as its means of expression (1988:94); furthermore, Hatim and Mason (1990:2) say "poetry is an act of self expression and not of communication".

Regarding these remarks, we can conclude that there may be all kinds of constraints and problems in the translation of poetry as a cultural heritage of one nation full of cultural terms. Totally, in poetry every word is a symbol which represents an area of experience or part of one's environment. So it requires the cultural competence of a translator in order to overcome the problems originating from culture- specific terms in the translation of poetry.

3-The Role of Interpretation: Translator's influence of the poem he reads affects the output. In other words, the translator's knowledge about the poem and his understanding of it can be considered as the most important factor in translating poetry. It may be proper to say that some poems need to be interpreted in order to be more understandable for the target language readers. So literal translation in some cases leads into bad results. But interpretation should not be so far from the main point presented by poet. The translator should not change the poem in his own taste. He have to be faithful to the source text, however, he may be authorized to write valid and reliable interpretation at times.

All of factors mentioned above indicate that translator should have the following competences: language, culture, subject, text and transfer one.

Here it is proper to move on Edward FitzGerald and his approach to translation. He was born to a distinguished family of Irish heritage on March 37, 1809 at Bred Field, near Woodbridge, in the Suffolk area of England. In 1830, he was graduated from Trinity College, Cambridge. He was a prolific letter- writer, corresponding regularly with such close friends as William Makepeace Thackeray and Alfred, Lord Tennyson and Thomas Carlyle. Beginning in the 1830s, FitzGerald wrote numerous poems and essays, nearly all of which went unpublished. His *Euphranor*, a philosophical dialogue, was published in 1857, followed by *Polonius*, an anthology of quotations in 1852. FitzGerald began studying Spanish ballads and drama, and his translation of six dramas by Pedro Calderon de la Barca was published in 1853. At the encouragement of his friend Edward Byles Cowell, a multilingual scholar, FitzGerald began learning Persian. In 1856 FitzGerald's translation of Jami's allegory *Salaman and Absal* was published. That summer, Cowell, just before

moving with his wife to Calcutta, India, discovered a manuscript of Omar's *Rubaiyat* at the Bodleian library at Oxford. At the time, Omar was known in his native land as a brilliant scientist and, somewhat less, so, as a poet. Cowell copied the manuscript to show FitzGerald. FitzGerald's *Rubaiyat*, translated anonymously and published in 1859, went virtually unnoticed until discovered by Dante Gabriel Rossetti the following Year. As its success modestly grew, FitzGerald felt compelled to revise it three times; these editions were published in 1868, 1872, and 1879. By the third edition, he was known to be the anonymous translator. FitzGerald had begun to translate Greek drama; his *Agamemnon* was published in 1865 and his version of *Oedipus Rex* and *Oedipus at colonus* appeared in 1880-81. But he is known primarily for his *Rubaiyat*, which became immensely popular in the late nineteenth century, after FitzGerald's death in 1883.

Results

His overall approach to translation: FitzGerald was attracted by the idea of genuine imitation being achieved by an accidental imitator, a writer who has not set imitation as a primary goal. Recognizing his own limits as a translator, and convinced of the severe limitations of translation as an enterprise, he nurtured a vision of good translation as imperfect re-creation in his translations and the liberties he took in translation served this ideal. He translated many literary works and he resorted essentially to the same approach in most of his translations, both eastern and western, preferring loose or very loose paraphrase to literal faithfulness. The approach is evident in his six Dramas of Caldron published six years before the first edition of the *Rubaiyat*. In recognition of the great liberties he took with *Aeschylus*, FitzGerald attached to his *Agamemnon* a preface justifying his translation practice, in which he argues that an extraordinary liberal approach offered the only hope he had of recreating the spirit of the Greek original. In the preface he refers to the scheme by which John Dryden classified translations according to degree of literal faithfulness: metaphrase (word-for-word translation), paraphrase (looser translation) and imitation; for him, loose rendering was the only rendering. He rejects to produce a more strictly faithful translation since he believes a literal version would scarce be intelligible. He wants a rendering congenial to readers, and, to create one, he has taken bold but necessary liberties. He believes that at all cost, a

thing must live; better a live sparrow than a stuffed eagle; a reader, not a scholar one, he aimed to please an interested but unscholarly reader. Totally he applied free and reader- oriented approach to all of his translated works. Now we are going to explore this approach and other possible strategies of translation in one of his translations: Khayyam Rubaiyat.

The tradition of translating the Rubaiyat of Khayyam

Omar Khayyam (1053-1123) was a Persian poet, astronomer, and mathematician whose poems are more widely known to English readers through Edwards FitzGerald's brilliant nineteenth century translations (1859).

Depending on the sources of reference that one chooses, Omar Khayyam is believed to have composed somewhere between 200 and 600 Rubaiyat. Some are known to be authentic and are attributed to him, while others seem to be combinations or corruption of his poetry, and whose origins are more dubious. It has been translated into most languages including English, French, German, Italian, Russian, Chinese, Hindi, Arabic and Urdu. In the following we list some of its translations into European languages. The main point is that none of the following translations obtained the popularity of FitzGerald's translation among the public.

Graf von Shack Adolf Friedrich von Shack (1815-1894) published a German translation in 1878.

Friedrich von Bodenstedt

Bodenstedt (1819-1892) published a German translation in 1881. The translation eventually consisted of 395 quatrains.

Edward Henry Winfield

Two English editions by Winfield (1836-?) consisted of 253 quatrains in 1882 and 500 in 1883.

J.B. Nicolas

The first French translation, of 464 quatrains in prose, was made by J.B. Nicolas, chief interpreter at the French Embassy in Persia in 1867.

John Leslie Garner

An English translation of 152 quatrains, published in 1888.

Justin Huntley McCarthy

McCarthy published prose translations of 466 quatrains in 1888

Richard Le Gallienne

Gallienne produced a verse translation, subtitled "a paraphrase from several literal translations", in 1897. In his introductory note to the reader, Le Gallienne cites MacCarty's "charming prose" as the chief influence on his version.

Edward Heron-Allen

Edward Heron-Allen (1861-1943) published a prose translation in 1898. He also wrote an introduction to an edition of Frederic Rolfe (Baron Corvo)'s translation into English of Nicolas's French translation.

Franz Toussaint

The best-known version in French is the free verse edition by Franz Toussaint (1879-1955) published in 1924. This translation consisting of 170 quatrains was done from the original Persian text, while most of the other French translations were themselves translations of FitzGerald's work. The *Éditions d'art Henri Piazza* published the book almost unchanged between 1924 and 1979. Toussaint's translation has served as the basis of subsequent translations into other languages, but Toussaint did not live to witness the influence his translation has had.

A. J. Arberry

In 1959, Professor A. J. Arberry, a distinguished scholar of Persian and Arabic, attempted to produce a scholarly edition of Khayyam, based on thirteenth-century manuscripts. However, his manuscripts were subsequently exposed as twentieth-century forgeries.

Robert Graves and Omar Ali-Shah

While Arberry's work had been misguided, it was published in good faith. The 1967 translation of the Rubáiyat by Robert Graves and Omar-Ali Shah, however, created a scandal. The authors claimed it was based on a twelfth-century manuscript located in Afghanistan, where it was allegedly utilized as a Sufi teaching document. But the manuscript was never produced, and British experts in Persian literature were easily able to prove that the translation was in fact based on Edward Heron Allen's analysis of possible sources for FitzGerald's work.

Peter Avery and John Heath-Stubbs

A modern version of 235 quatrains, claiming to be "As literal an English version of the Persian originals as readability and intelligibility permit", published in 1979.

Karim Emami

In 1988, for the very first time the Rubaiyat were translated by a Persian translator. Karim Emami translated the Rubaiyat in his title "The Wine of Nishapour" which was published in Paris. The Wine of Nishapour is the collection of Khayyam's poetry by Shahrokh Golestan.

Ahmed Rami

Ahmad Rami, a famous late Egyptian poet, translated the work into Arabic. His translation is considered to be a most fascinating work of modern Arabic literature.

Due to the great reception of FitzGerald's translation of Rubaiyat, now we want to explore his approach in its translation based on these factors: form , meaning ,interpretation of poem , and culture – specific terms.

In terms of stylistic and formal features like rhyme, alliteration or other kind of figures of speech, he reproduced a translated poem that seems to be an original one so that it is claimed that FitzGerald was better at rearranging the creations of others than he was at inventing his own. Success in translation meant, for FitzGerald, the re-creation of a poetic voice .Regarding this fact and his main career as poet, he is more successful in recreating stylistic features. From this point, English readers appreciate him and call his translation as Omar-Fitz poems. But there is a basic deficit in his translation; Khayyam's Rubaiyat are originally arranged according to alphabetical rhyme, a Persian tradition of arranging sonnets. But FitzGerald arrange them thematically in a way that Khayyam enters a Persian garden at dawn , conscious and alert, then sinks into contemplation, drinks during the days, gradually attains to a state of exhilaration, writes his poetry to celebrate the glory of the garden , and finally mourns at man's short life as night approaches .

Meaning and the way he interpreted Rubaiyat.

In his translation, the sequence of a day acts as a metaphor for the passage of time. It extols the hedonistic pleasures of food, sex and wine, and the importance of living for today, because the future is uncertain and life is fleeting. Written during a time of religious upheaval its first edition was published the same year as Darwin's origin of species – the translated poems questioning of religion and traditional morality was both shocking and fascinating to its reader. So FitzGerald interpreted the Rubaiyat on his own taste and focused on the literal meaning. FitzGerald assigned an Epicurean interpretation to the Rubaiyat and interprets them in a way to fit his translation into

the western frame of thought and pain himself fame he would perhaps never achieve otherwise:
the best example is this Rubai:

Ah! My Beloved, fill the cup that clears

To- day of past Regrets and Future Fears

اي دوست بيا تا غم فردا نخوريم
وين يکدم عمر را غنيمت
شمرم

Culture- specific Terms

To translate these terms, he used different strategies. In some cases, he used borrowing strategy and imported that name. It is more significant in translating proper names like کيقباد و رستم و حاتم طاعی و کيخسرو و بهرام و جمشيد.

By using this strategy, he maintained the oriental flavor. Another strategy he used is domestication; this is obvious in this example:

The Rose as where some **buried Caesar** bled:

آن لاله ز خون شهرياری بودست

This is not full equivalent of Persian term. Their connotation is different. The point is that Caesar was killed by an Iranian and it is not appropriate to use as an equivalent. The equivalences such as beloved, Tavern, Heavenly Master, lovely houris are not the total and real equivalents of the terms
كوزه و فلک و ميخانه و دوست

Regarding the abovementioned aspects on the translation of Khayyam Rubaiyat, we can classify translation strategies in the following way:

1. Addition : According to Heron Allen , forty four of FitzGerald's quatrains are translated and composed of only one or two quatrains in Khayyam's : This means that these are not translation , but FitzGerald himself composed original English poems based on part of Khayyam Rubaiyat such as the following quatrain from FitzGerald's translation:

Then of the thee in Me who works behind

The veil , I lifted up my hands to find
A lamp amid the darkness ; and I heard,
As from without. (The Me within Thee blind)

For this Rubai, there is no equivalent in Persian original source.

2. Selection: FitzGerald prepared his own version based on Bodleian and Calcutta version consisting of 158 and 516 Rubai respectively. He made this selection probably due to these factors: how much he found them translatable, how much they fit into his Epicurean interpretation, etc.
3. Omission: he omitted some quatrains either because he found them difficult to translate or because they didn't conform to the Epicurean image he had constructed of Khayyam.

Example:

And this I know: whether the one True light,
Kindle to Love, or Wrath consume me quite,
One Glimpse of it within the Tavern
Better than in the Temple lost outright

به زانکه بحراب کنم راز و	با تو به خرابات اگر گویم راز
	نیاز
خواهی تو مرا بسوز و خواهی	ای اول و ای آخر خلقان همه تو
	بنواز

4. Domestication: In some instances FitzGerald used domestication strategy proposed by Venuti and acculturates Persian terms.

Example:

I sometimes think that never blows so red
The Rose as where some **burid Caesar** bled;
That every Hyacinth the Garden wears
Drop in its Lap From some one lovely Head

هر جا که گلی و لاله زاری می بودست آن لاله ز خون **شهریاری** بودست

هر برگ بنفشه کز زمین می روید خالیست که بر روی نگاری بودست

5. Foreignization: it is another term as a strategy in translation proposed by Venuti. Here FitzGerald imports the Persian terms into English by transliteration and foreignize his translation. This is more obvious in the case of culture – specific terms.

Example:

They say the lion and the lizard keep

The courts where **Jamshyd** gloried and drank deep:

And **Bahram**, that great Hunter-the wild Ass

Stamps Oer his Head- and he lies fast sleep.

آن قصر که جمشید در او جام گرفت آهو بچه کرد و رو به آرام گرفت

بهرام که گور می گرفتش همه عمر دیدی که چگونه گور بهرام گرفت

Totally, he used all of possible strategies to produce a free translation of Khayyam Rubaiyat to be accepted greatly among the English public. In the following section we explore the possible reasons for these strategies.

Discussion

The main purpose of this study was to find out and classify theories and strategies used in translation by FitzGerald especially in the translation of Khayyam Rubaiyat. The theory he used – he mentioned it in the preface of *Agamemnon*- was the scheme by which John Dryden classified translations according to degree of literal faithfulness: metaphor (word – for – word translation), paraphrase and imitation; he used paraphrase and sometimes imitation as a translation theory in translating. He held this attitude in all his translated works. In order to realize this theory, he used some strategies .In the case of Khayyam Rubaiyat, he used mainly these strategies; addition, selection, omission, domestication, foreignization. These strategies are found out based on the analysis of translation in terms of form, meaning, interpretation, and culture - specific terms. It seems that FitzGerald exercised this approach (free translation) and strategies mentioned above because of the following reasons:

- Limited knowledge of the Persian language.

- Ignorance of the tradition of Persian poetry.
- The marginal position he assigned to Persian literature.
- Using translation as a mode of writing and composing poems.
- Producing a very natural translation to be pleasant for all English readers.

FitzGerald's knowledge of the Persian language was extremely poor and his resources for the Persian language consisted of sir William Jones Grammar of the Persian language and a dictionary of Persian, Arabic and English and also his familiarity with Cowell who knew Persian. So Fitzgerald translated the Rubaiyat with such limited knowledge and resources; therefore he was not cognizant of the tradition of Persian poetry. Also, FitzGerald, as a result of the colonial attitude dominant in his time, considered Persian literature inferior as compared to the superior position he attributed to English literature. FitzGerald's disparagement of Persian literature involved prejudiced views that were common in Victorian Britain and that have a place in orientalism. He held ideas that the poetry could be understood with an incomplete knowledge of Persian, that it would benefit from European rewriting, that it was minor literature, "little" and childishly devoted to simplistic and repetitive motifs. Fitzgerald did not confine his efforts to improve literature to Persian poetry alone. He liberally cut and revised the poems of George Crabbe. The liberties he took with verse of the Quaker poet Bernard Barton also resemble those he took with Khayyam. Fitzgerald believed that certainly he had bettered the original. It seems that having these ideas allowed him to exercise any kind of liberty in translation.

The main limitation in this study was that I as writer only know English and Persian. If I knew the Greek language, it would be possible to assemble some examples from the translation of the Greek texts by FitzGerald and certainly it would lead to better results about his overall approach. Whatever I mentioned here about his overall approach to translation was borrowed from the English researches on this topic. In this study I only focused on Khayyam Rubaiyat as an instance of FitzGerald translated works. It would be

useful to discuss and explore his overall approach in translation in other translated works of FitzGerald and to compare them to the translation of Khayyam Rubaiyat.

Conclusion

As it was mentioned, translation plays a crucial role in the construction of history. In other words most of the world's past comes to us in translation and it is reasonable to say that the history of the world could be told through the history of translation. So, through translation, people have gained opportunity to become familiar with the other nations' literature. In order to translate a text, translators adhere to different approaches which subsequently affect its reception and popularity in the TL. In this paper we attempted to explore FitzGerald's approach to translation, especially his approach to the translation of Khayyam Rubaiyat and tried to find the reasons underlying his basic approach in translation. Through this examination, it become evident that there is a direct link between his theory and practice and also that Fitzgerald used appropriate methods including selection, omission, foreignization and domestication to realize his ideal translation. Since every research is incomplete and continuous process and this study only serves a partial part of this study, there are many intact areas to do more specific researches in this respect.

References

- Bassnet,S.(1992).*Translation Studies*.Routledge:London&New York
- Drury,A.(2008).*Accident,Orientalism,and Edward FitzGerald as Translator*.Retrieved 2008/11/24 from <http://www.freearticles.com>.
- Emami, K.(1375). *Az Past va Bolande Tarjome*. Nilufar Publishers: Tehran
- Gray,E.(2008).*Fitzgerald and Rubaiyat,in and out of time*.Retrieved 2008/11/24 from <http://www.freearticles.com>.
- Gray,E.(2008).*Victorian Poetry*.Retrieved2008/11/24 from <http://www.freearticles.com>.
- Hatim,B&,Mason,I.(1990).*Discourse&The Translator*.Longman:London&New York
- Lazim,H.G(2007).*Poetry Translation*. Retrieved

from www.translationdirectory.com/article1362.htm.

Munday, J. (2001). *Introducing Translation studies*.

Theories & applications. Routledge: London & New York.

Newmark, P. (1998). *A Textbook of translation*. Prenticehall: Singapore.

Nida, E. A. (1964). principles of correspondence in: L. Venuti. *The Translation Studies Reader*. London & New York: Routledge. pp. 126-140

Venuti, L. (2000). *The Translation Studies Reader*. Routledge:
London & New York.

<http://en.Wikipedia.org/wiki/FitzGerald>

¹ . Habibollah Mashhady, Assistant Professor, Department of English Language , University of Zabol, Mashhadyh@yahoo.com

² . Mahbube Noura, Faculty Member of English Language Department, University of Zabol,
nouramahbube@yahoo.com

Investigate to Substitute Large Inertia by a Combination of Very Small Inertia Driven by v/f Controlled Drive

M. V. Palandurkar, J. P. Modak, S. G Tarnekar

Abstract — Process machines for fluctuating load torque use large flywheels with high moment of inertia. They suffer from the disadvantages of bulky systems prone to reduced acceleration. As a result these are harder to start and are likely to have frequent mechanical breakdowns. A method is suggested here to effectively work with much smaller flywheels. Suitably monitored VVVF drives and low moment of inertia offer a much better alternative. The system behavior is drastically improved. This paper deals with effective energy transactions in conventional systems to compare the same in case of the proposed system. The method is verified here with MATLAB simulation, based on which it is reported that good choice of frequency to suit the nature of load torque fluctuations resulted into optimum system performance

Index Terms— Process Machine, Flywheel, Demand Torque, Induction motor, *v/f* Control, moment of inertia, Modeling, simulation.

1 INTRODUCTION

Generally there are two types of process machines, Type I and Type II. In Type I process machine, input motion is completely rotary and output motion is also rotary at uniform speed. In such process machine, demand torque does not vary cyclically, but varies with the load. Type II process machine makes use of link mechanism or cam mechanism or combination of linkage, cam and gears. For such process unit, at every instant, demand torque changes with respect to time. The arbitrary demand torque characteristics of any process machine can be estimated based on cycle time of operation, process resistance and inertia resistance [1], [2]. Hence, this variation is cyclic and cycle time is commensurate with rpm of process unit. But, usually, induction motor cannot generate closely matching torque characteristics. Hence, the flywheel is required to make up for the difference of the torque in different time intervals. Fig.1 describes an arbitrary demand torque characteristic of a process machine, where, T_d is demand torque and T_s is average of the electromagnetic torque generated. This can be estimated based on cycle time of operation, process resistance and inertia resistance. These can be detailed based on intended operation and proposed details of partial mechanical design [1, 2]. It is evident from Fig. 1 that crank speed of input shaft of the process machine should be 20 rpm. Time for complete cycle of operation of the process unit should be 3000 msec which gets completed in one rotation of the input link of the process machine.

Fig. 1 shows that demand torque has fast variation with time, which the motor cannot cope up with this. Hence, the flywheel is required to make up for the difference of the torque in all sections of time axis. It is known that flywheel will decelerate during intervals AB & CD when load torque is greater than average electromagnetic motor torque whereas it will gain speed during intervals O'A, BC, DE sections of time axis when load torque is less than average electromagnetic torque.

Fig. 2 describes the schematics of an arbitrary process unit P along with usual mechanical power transmission system for torque amplification and speed reduction. Pulley D2 is a flywheel cum power transmission pulley, pulley D1 is driving pulley which receives power from the Induction motor M. The portion of the system between D2 and process unit is subjected to severe torsional vibrations. Hence, presence of flywheel with high moment of inertia J in the process machine increases power rating of the main drive, reduced acceleration, increased weight of engine, harder to start and fatigue in the component of power transmission system [3]. Therefore, it is desirable to eliminate flywheel from the design of any process machine in general. It is felt that by properly designing power electronic circuitry using VVVF drive having low moment of inertia, this may be possible. Hence, in the present paper, among different control schemes, a constant volt per hertz, principle is chosen to drive three phase induction motor as

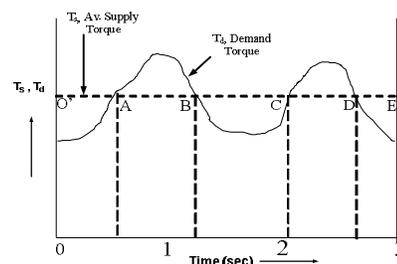


Fig.1. Arbitrary Demand Torque Characteristics.

- M. V. Palandurkar is currently pursuing Ph. D degree program in Electrical and Mechanical Engineering in GHRCE and Asst. Professor in RCEOM, both under RTMNU University, Nagpur, MS, India, PH- +919823941014. E-mail: mpalandurkar@yahoo.com
- J. P. Modak, Emeritus Professor, Mechanical Engineering Dept. PCE, Nagpur, MS, India, PH- +919890831037. E-mail: jpmadak@yahoo.com
- S. G. Tarnekar, Professor, Electrical Engineering Dept. GHRCE, Nagpur, MS, India, PH- +919890831037. E-mail: sgtarnekar@gmail.com

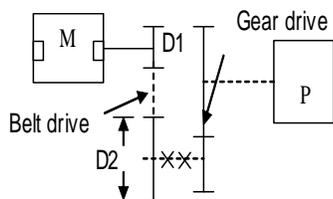


Fig. 2. Schematics of an Arbitrary Process Unit, Mechanical Power Transmission & 3 phase Induction Motor.

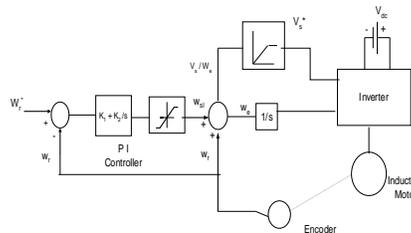


Fig. 3. Induction Motor Drive with Closed Loop Volts / Hertz Control.

shown in Fig. 3. In this technique, a dynamic model of three phase induction machine is derived from two phase machine. [4], [5]

The equivalence between three phase and two phase machine is based on the equality of the mmf produced by two phase winding and by three phase. The stator and rotor variables are transformed to a synchronously rotating reference frame that moves with the rotating magnetic fields. Finally, a dynamic machine model in synchronously rotating and stationary reference frame is developed in per unit by defining the base variables both in $a-b-c$ and $d-q-o$ variables. Authors have already reported [9], that the above method can be analyzed. It uses VVVF based induction motor drive by controlling input side frequency for better performance, with much smaller system inertia. According to change in demand torque, varying cyclically with respect to time, the requirement of input frequencies to the main drive during different time intervals also changes in order to generate electromagnetic torque characteristics matching with demand torque characteristics. Hence problems occurring due to the presence of large flywheel between induction motor and process machine are eliminated.

Generating electromagnetic torque to match demand torque characteristics of the process machine by the above method is not only a task to fulfill but it is also necessary to find how much energy is transacted from motor side when demand load torque sudden changes low to peak value. Hence, the present paper compares the effective energy transaction in conventional system having large flywheel with VVVF based induction motor drive of low moment of inertia of much smaller flywheel.

Induction motor coupled with conventional flywheel having large moment of inertia is run at constant frequency without applying any control technique whereas, in lateral case, induction motor is chosen to have low moment of inertia and is controlled by input side frequency using VVVF technique. The effective energy is calculated in both the case by using kinetic energy formula.

If rotor is moving at mechanical speed w_m with moment of inertia J , the kinetic energy produced is given by the formula,

$$K.E = \frac{1}{2} J w_m^2 \tag{1}$$

Differentiating with respect to time, we get power, i.e.

$$\frac{d(K.E.)}{dt} = J w_m \frac{dw_m}{dt} \tag{2}$$

Equation (2) enables us to calculate the energy transacted for a particular case. This has been the basis of comparing energy delivered by the rotating masses to the shaft during first dip in speed. This enables to conclude about the proposed scheme.

2 CLOSE LOOP INDUCTION MOTOR DRIVE WITH CONSTANT VOLTS PER HERTZ CONTROL STRATEGY

An implementation of the constant volts / hertz control strategy for the PWM inverter fed induction motor with suitable value of moment of inertia in per unit basis is simulated in MATLAB simulink software is shown in Fig. 4 [6]. The major blocks consist of PWM inverter, induction motor with mechanical load [7], [8] [12]. In this scheme mechanical load varying cyclically in an assumed pattern as shown in Fig. 5. As the load torque increases, the speed loop error generates the slip speed command w_{sl} through proportional-integral controller and limiter. The slip is added to the speed feedback signal w_o to generate the slip frequency command w_e . The slip frequency command generates the voltage command, V through a volts/hertz function generator. A step increase in slip fre-

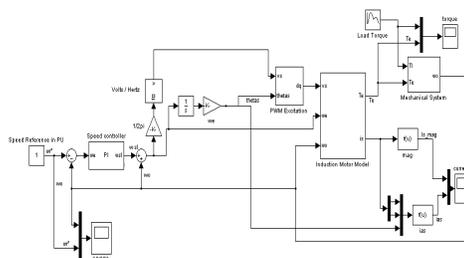


Fig. 4. Complete Induction Motor Model with PWM Excitation and Mechanical System along with v / f control scheme in MATLAB SIMULINK.

quency command produces a positive speed error and the slip speed is set at the maximum value. The drive accelerates due to changes in the frequency and current, producing the torque, matching with demand torque characteristics. The drives finally settle at a speed for which motor torque equals the load torque. Hence, for varying load torque with respect to time,

TABLE 1
 INDUCTION MOTOR DATA

the drive generates similar electromagnetic torque which matches with demand torque of the process machine.

3 CASE STUDY

A process machine under consideration comprises of a linkage mechanism as a main processor. On account of non linear kinematics of the mechanical hardware, demand torque characteristics of the process machine are time-variant. In an attempt to simplify the analysis, and tests the proposed system, using standard logics in control system engineering, the demand torque variation is as shown in Fig. 5. At t= 1.5 sec, the torque suddenly rises by a step. Sudden rise in torque needs extra kinetic energy to be fed to the shaft. Deceleration is the result. And first dip in speed contributes maximum to counter the sudden rise in torque. Therefore, first dip is expected to be most crucial for energy calculations. Here speed of the process machine is chosen to be 20 rpm, hence the total cycle time is 3000 msec. The induction motor rating is 3 phase, 415 V, 1 HP with a synchronous speed as 1500 rpm. In this case, the average angular velocity of the input crank of the process unit must be 20 rpm. This gives torque amplification from motor shaft to the process unit input shaft of the order of 75. This induction motor generates average supply torque of 5.96 N-m (with given torque formula [4]) or 0.596 Kgf m. Thus the supply torque at the process unit input shaft is 0.596 X 75 = 44.7 Kgf-m. Hence, the hp demand of the process unit with a given formula is,

$$\begin{aligned}
 hp &= \frac{2 \cdot \pi \cdot N \cdot T}{4500} \\
 &= \frac{2 \cdot \pi \cdot 20 \cdot (0.596 \times 75)}{4500} \\
 &\approx 1.248
 \end{aligned}$$

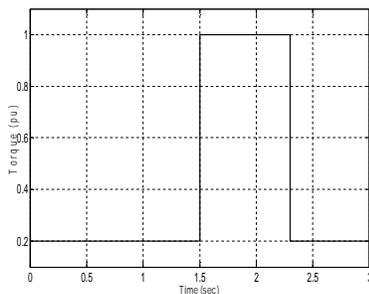


Fig. 5. Demand Torque characteristic of a specific Process Machine.

4 SIMULATION AND RESULT

In order to get the results, induction motor with two different moments of inertia is considered for simulation. In first case

HP	1 = 0.75 kW
Rated Voltage	415 V, ± 10% tolerance
Winding Connection	Star
Rated Frequency	50Hz
Pair of poles	2
Rated speed	1500 rpm
Stator Resistance	12.5487 Ω
Rotor Resistance	12 Ω
Stator Leakage Inductance	144.67 mH
Rotor Leakage Inductance	144.67 mH
Mutual Inductance	545.78 mH
Moment of Inertia	0.0018 kg m ²
Friction Factor	0.01

induction motor with high moment of inertia is chosen to drive it at constant frequency. In second case induction motor with low moment of inertia is considered by controlling input side frequency using close loop VVVF drive. Here, all parameters of induction motor model are converted into per unit basis [5] and the main drive is simulated for one complete cycle of rotation of 3000 m sec to get the results.

After simulation of operation at constant frequency and with high moment of inertia, it is observed that induction motor generates similar nature of electromagnetic torque with respect to demand torque. Motor torque matches its load torque characteristics as shown in Fig. 6. Fig. 7 shows the simulation of induction motor by monitoring its input frequency using VVVF drive with low moment of inertia. This simulation shows that the electromagnetic torque is almost matching with demand torque characteristics requiring less starting period as compared to previous case. [9] When load torque suddenly increases to high value, generated electromagnetic torque is less than load torque during transient period hence, speed decreases. But when load torque decreases to low value, generated torque is greater than load torque in transient period hence, speed increases. Fig. 6.1 and Fig. 6.2 show decrease and increase of speed when load torque increases to peak value i.e. 1 pu and decreases to 0.2 pu respectively. Drop in speed with constant input frequency and with high moment of inertia of J = 0.018 kg-m² is 0.9174 pu ie. 144.10 rad/sec with settling time of 0.35 sec. and increase in speed is 0.9974 pu i.e. 156.67 rad/sec with settling time of 0.3 sec. Here, base mechanical speed is considered as 157.079 rad/sec. The amount of effective energy is required to generate similar type of demand torque of process machine is 321.71 Joules.

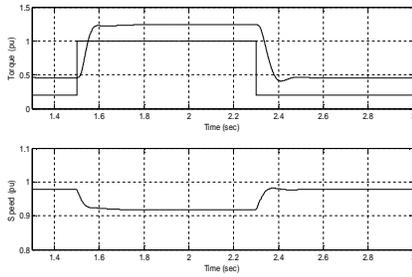


Fig. 6. Electromagnetic torque and speed of induction motor for load torque without VVf with high value of $J = 0.018 \text{ kg-m}^2$

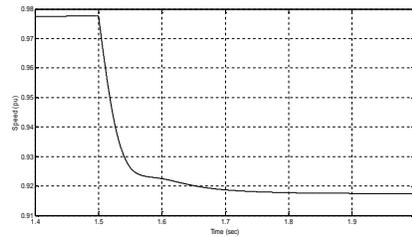


Fig. 6.1. Drop in speed for increase peak value of torque of $J = 0.018 \text{ kg-m}^2$.

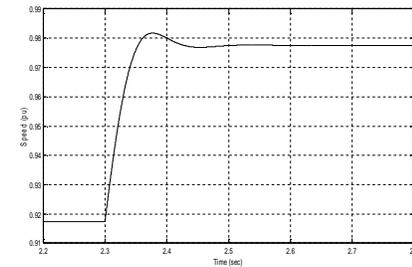


Fig. 6.2. Rise in speed for decrease of torque of $J = 0.018 \text{ kg-m}^2$

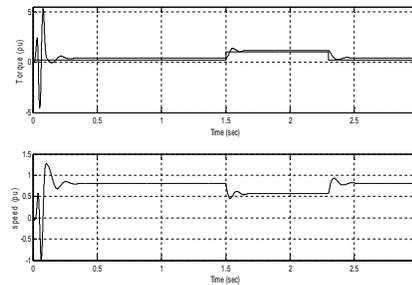


Fig. 7. Electromagnetic torque and speed of induction motor for load torque VVf with low value of $J = 0.0018 \text{ kg-m}^2$

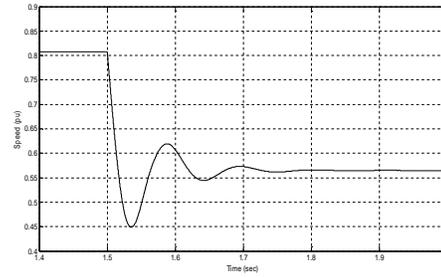


Fig. 7.1. Drop in speed for increase peak value of torque of $J = 0.0018 \text{ kg-m}^2$.

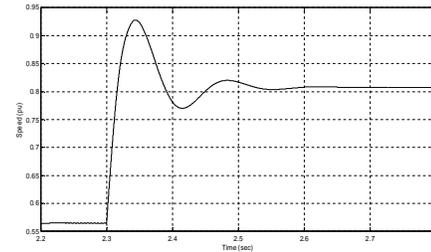


Fig. 7.2. Rise in speed for decrease of torque of $J = 0.0018 \text{ kg-m}^2$.

In second case, by controlling input frequency of using VVf drive with low moment of inertia of $J = 0.0018 \text{ kg-m}^2$, electromagnetic torque is generated at reduced frequency 32.61 Hz to match the demand torque characteristics with less error. The drop in speed due to sudden increase in torque is more as compared to first case and it is observed as 0.565 pu i.e. 88.75 rad./sec with same settling time i.e. 0.34 sec. as shown in Fig. 7.1 The increase in speed due to decrease of torque is 0.8075 pu i.e. 126.84 rad./sec. with settling time 0.35 sec. as shown in Fig. 7.2. The required energy to produce similar type of load torque is 243.40 Joules which is less than that in previous case. Table 2 shows the calculation of energy due to sudden drop in speed when torque changes from low to high value.

5 CONCLUSION

In order to generate electromagnetic torque matching with demand torque characteristic of the process machine with less energy transaction, it is possible to run induction motor with much smaller flywheel using VVf drive suitably controlling its input frequency. Hence, larger flywheel with high moment of inertia can be replaced with smaller one. From the result, it is clear that require energy for sudden change of load torque to peak value is more during transient period when speed drops drastically in conventional system with larger flywheel when motor runs at constant frequency. Opposite to previous case, require effective energy transaction from rotational masses to shaft of the motor to match the change of load torque to peak value is less when drive is controlled from input side frequency with low moment of inertia using VVf technique, electromagnetic torque almost matches the demand torque at reduced frequency and acceleration of induction motor is also improved. Variation in speed is more oscillatory

TABLE 2

CALCULATION OF POWER FOR TWO DIFFERENT VALUES OF MOMENT OF INERTIA, J, BASE MECHANICAL SPEED, $W_m = 157.079$ RAD./SEC

Sr. No.	J (kg-m ²)	Fre-quency (Hz)	Time span (sec)		dt (sec)	Speed Range (pu)		d(W _m) (pu)	d(W _m) rad/sec	$\frac{d(W_m)}{dt}$	W _m (pu)	W _m (rad/ sec)	$J * W_m * \frac{d(W_m)}{dt}$ (Joules)
1	0.018	50	1.5	1.57	0.07	0.977	0.923	0.054	8.482	121.176	0.939	147.49	321.71
2	0.0018	32.61	1.5	1.537	0.037	0.806	0.449	0.357	56.077	151.61	0.568	89.22	243.40

and settling time of fall in speed from transient period to steady state is same for both the cases. Hence, depending on torque fluctuation and the inertia, the optimum frequency which resulting into extracting energy from rotational masses is work out and found that drives with low moment of inertia suitably monitored offer a much better alternative to large flywheel to generate same load torque variation. Behavior of system is drastically improved.

6 SCOPE OF FUTURE WORK

Having reported the basic confirmation (using MATLAB) of the proposal to eliminate the bulky flywheel, following aspects can be investigated in future work

1. Estimation of necessary moment of inertia in view of some difference in the required demand torque characteristics and obtained generated demand torque characteristics.
2. Estimation of initial additional investment by incorporation of electronic circuitry and corresponding mechanical transmission. Estimate in both ways in terms of initial investment and operational cost.

REFERENCES

[1] Joseph Edward Shigley and Charles R. Mischke, "Mechanical Engineering Design", 2nd edn, Tata McGraw Hill, Ed. (Book style)

[2] Joseph Edward Shigley and John Joseph Vicker Jr., "Theory of Machines & Mechanisms", 2nd edn, McGraw Hill international (1995) (Book style)

[3] James D and Ven de Ven, "Fluidic Variable Inertia Flywheel" American Institute of Aeronautics and Astronautics, pp.1 -6

[4] Bimal K. Bose, "Modern Power Electronics and AC Drives", 5th edn, Prentice Hall PTR, pp. 56-74, 264-269, 339-345, 2005 (Book style)

[5] R. Krishnan Bharat, "Electric Motor Drives, Modeling, Analysis and Control", PHI Learning pvt. Ltd, pp. 196-224, 2008. (Book style)

[6] Burak Ozpineci, Leon and M Tolbert, "Simulink Implementation of Induction Machine Model"- A Modular approach. IEEE Transaction on Industry Applications, pp. 728-734, 2003. (IEEE Transaction)

[7] Hussein Sarhann and Rateb Issa, "Improving Mechanical Characteristics of Inverter Induction Motor Drive System" American Journal of applied Sciences3, Vol..8, pp 1961-1966, 2006 (URL Journal)

[8] Ogbuka C. U and Nwosu C., "AGeneralised Vector Method of Induction Motor Transient and Steady State Analysis". The Pacific Journal of Science and Technology, Vol. 10(1), pp.52-58 2009 (URL Journal)

[9] Palandurkar M. V., Modak J. P. and Tarnekar S.G., "Elimination of a Flywheel of a Process Machine by Controlling Power Frequency of the Main Drive". International Journal of Engineering Science & Technology (IJEST), Vol. 3, pp. 3580-3590, 2011 (URL Journal)

[10] Alfredo Munoz, Garcy Thomas, Lipo, A, Fellow IEEE, Donald W and Novotny, Fellow IEEE, "A New Induction Motor V/F control Me-

thod Capable of High Performance Regulation at Low Speeds" IEEE Transaction on Industry Applications, Vol. 34 (4th July/August), pp.813-821 1998 (IEEE Transaction)

[11] Shit K. L, Chan T. F, Wong Y. K and Ho S. L, "Modeling and Simulation of the Three Phase Induction Motor using Simulink", International Journal of Electrical Engineering Education, Vol. 36 (Manchester U.P.), pp. 163-172 1999 (URL Journal)

[12] Jamesh and Nash N, Member IEEE, "Direct Torque Control, Vector Control without Encoder" IEEE Transaction on Industry Applications, Vol. 33, 2 March / April 1997 (IEEE Transaction)

[13] Noaman M. Noaman, "Speed Control for IFOC Induction Machine with Robust Sliding Mode Controller" Asian Journal of Scientific Research, Vol. 4, pp.324-333 2008 (URL Journal)

Overview of Optical Interconnect Technology

Sumita Mishra, Naresh K Chaudhary, Kalyan Singh

Abstract— Optical interconnect is seen as a potential solution to meet the performance requirements of current and future generation of data processors. Optical interconnects have negligible frequency dependent loss, low cross talk and high band width. Optical interconnects are not much used commercially since optical interconnects technology is incompatible with manufacturing processes and assembly methods that are currently used in the semiconductor industry. There are many promising optical interconnect technologies and this paper presents a brief analysis of current state of optical interconnect technology.

Index Terms— optical interconnect, optical source, Detector



1 INTRODUCTION

In recent years, the performance of single chip multiprocessor has been roughly doubling every 18 months resulting in tremendous increase in its clock speed and cache size moreover multiple cores and hyper threading has increased the complexity of these systems.

These multicore ultrafast processors require high speed interconnects that allow individual processors fast access to memory, its cores and other I/O devices. As speed and complexity of these systems increase the interconnect density and throughput management becomes a critical factor towards the realization of high performance data processing systems. Currently copper interconnects are used for data transmission over chip-to-chip and chip-to-module interfaces, in chip-to-chip over backplane, and in chip-to-chip over copper cable assemblies. Electrical interconnects performance degrades at frequencies above 1 GHz due to ringing, increased signal latency, crosstalk and frequency dependent attenuation [1]. These limitations of electrical interconnects will limit the maximum frequency of operation for future systems; Optical interconnect is seen as a potential solution since it can directly address these problems at the system level and meet the performance requirements of current and future generation of data processors, optical interconnects have negligible frequency dependent loss, low cross talk and high band width.

Despite the significant interest shown by many groups worldwide, optical interconnects are not much used commercially. In order to become a viable technology to replace electrical-based on-chip interconnects, optical interconnects should be made compatible with manufacturing processes and assembly methods that are already in use in the semiconductor industry further there is a need to develop efficient and compact optical interconnect modules that use simple optical and electrical interfacing schemes. The development of optical interconnects, especially based on a technology platform which is monolithically integratable into Si CMOS at low cost is needed in order to make optical interconnects economically viable. This will result in low cost, high performance and CMOS compatible optical components. Since it is not possible to make silicon light emitting and detecting we have to integrate other materials with Si. Large-scale integration of optical devices has been demonstrated on III-V platforms but in this implementation the components have different technology and they cannot be monolithically integrated on the same substrate. Monolithic integration of optical and electronic components on one substrate [2-7] together with demonstrations of efficient fibre to waveguide couplers [8] has shown the promise for development of ultra-compact optical components compatible with current technology.

2 BASIC COMPONENTS

The main components of an optical interconnect system are shown in Figure 1. The optical modulator has two inputs, optical signal from the off chip laser source and the electrical signal from the CMOS driver

Sumita Mishra is currently pursuing doctoral degree in Electronics at DRML Awadh University, India
E mail: mishra.sumita@gmail.com

circuit. Optical couplers are structures that are used to inject the light into the optical system. Electrical signals that are to be transmitted to some destination in an optical interconnect system must be converted into the optical domain for transmission. Modulator converts the electrical signal into optical signal according to bit sequence in electrical signal .After the optical information signal has been generated, it is fed into the optical routing structure. Optical interconnects use waveguides for signal transmission, which consists of dielectric materials with high index of refraction surrounded by a material with lower refractive index. Optical switches are used in optical routing networks to route the light travelling in waveguides to different locations.

The receiver side of the optical interconnect system is responsible for reconstruction of electrical signal. A photo detector is the device for detecting the light pulses and converting them to photo current. A trans-impedance amplifier is finally used for amplifying the photo current and providing the digital signal in the form of conventional voltage signal.

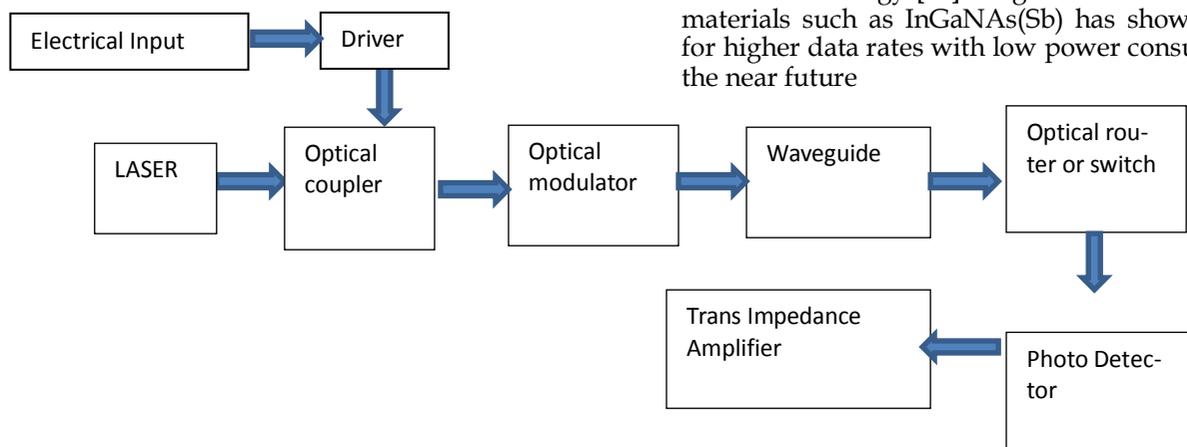


Fig 1. Optical interconnect block diagram

2.1 Laser source

Vertical Cavity Surface Emitting Laser (VCSEL) is currently the most attractive optical source for the short-distance optical interconnects. A Vertical Cavity Surface-Emitting Laser (VCSEL) is a semiconductor laser diode that emits light perpendicular to the upper surface of the semiconductor wafer of which the laser is composed. Monolithic arrays of high density VCSELs [ref] have been developed by various manufacturers which are

competitively priced. VCSELs can be manufactured for several different wavelengths [9]. Today, devices emitting at around 850 nm represent the most mature high speed-optimised VCSEL technology these devices are top emitting due to large substrate. VCSELs operating at 980 nm are based on GaAs/AlGaAs-based process technology. These devices have advantage of emission through the substrate, since GaAs is transparent at 980 nm. This brings added benefits such as integrating backside micro lenses for improved coupling [10]. 980 nm VCSELs have been the wavelength of choice for an ultra-dense chip-to-chip level interconnects, operating at data rates in the range of Tb/s . During the recent years there has been a significant development effort towards VCSELs operating at a wavelength range of 1200...1600 nm using the InP-based process technology. VCSELs operating at a wavelength near 1550 nm have the disadvantage of growth difficulties and low thermal conductivity of the necessary DBRs. [11] VCSELs offer many of the desirable characteristics of an optoelectronic transmitter for optical interconnects. 2D arrays of high speed VCSEL with good contrast at a low voltage drive can be fabricated . Recent advances in VCSEL technology [12] along with research into new materials such as InGaNaNs(Sb) has shown promise for higher data rates with low power consumption in the near future

2.2 Optical modulator

Optical interconnect systems may use either off-chip or on-chip laser sources. Off-chip laser sources are generally used due to several problems faced by directly modulating Lasers. Due to non-availability of an efficient silicon based laser it is very hard to fabricate large no of lasers on a single chip at reduced cost further it has added disadvantage of complex chip design

since optical source is part of the chip's power and heat budget. In order to modulate lasers at high bit rates, they must be operated well above threshold in both 0 and 1 bit states, which consumes high power and leads to performance degradation. One of the critical issues of future systems is to keep power budgets manageable while increasing performance this is achieved by increasing parallelism. Wavelength division multiplexing offers massive parallelism. WDM components are not an essential part for optical interconnect systems but they are considered very important in order to build high performance optical interconnects. The heat generation from on chip lasers is undesired since the temperature variation in CMOS chips also causes wavelength shifts and this instability can prohibit precise channel allocations for wavelength Division multiplexing in the same medium. So it is preferable to use on-chip modulators for transmitters and modulate the light coming from an off-chip continuous-wave (CW) laser. The design of a fast and cost efficient CMOS compatible electro-optical modulator is one of the most challenging tasks towards realizing on-chip optical interconnects.

There are various optical modulation techniques through which refractive index or absorption properties of optical medium are varied in accordance with the electrical signal; this variation in optical property of medium causes phase or amplitude modulation of optical signal. Current optical modulation techniques are based on Thermo Optic effect, Electro optic effect,[13] Electro absorption effect and plasma dispersion effect. Thermo-optic effect is the change in optical absorption coefficient and refractive index of an optical medium due to the change in temperature of the medium. Refractive index variation due to applied electric field is called Electro Optic Effect. The linear change in a optical medium's refractive index by an electric field is called the Pockels effect and if the change is quadratic, it is called the Kerr effect. Another modulation mechanism known as the electroabsorption effect involves the change in the absorption coefficient of the material with change in applied electric field. The Franz-Keldysh effect [14] is electroabsorption observed in bulk semiconductors and the Quantum Confined Stark Effect [15] is electroabsorption seen in quantum-confined structures such as thin quantum-well layers.

These modulation techniques are quite effective in III – IV semiconductors but in silicon refractive index change produced by these effects is very small since unstrained pure crystalline silicon does not exhibit Pockel's effect, Franz-Keldysh effect and the Kerr ef-

fect are very weak in Si Therefore; very high electro-magnetic field is required in order to achieve a useful change in the refractive index. The most effective mechanism for changing the refractive index in Si is the carrier plasma dispersion effect [16] in which the concentration of free carriers in silicon changes the refractive index and the optical absorption.

The modulator can be either refractive or absorptive type. Refractive modulators employ either single interference Mach-Zehnder interferometer structure or multiple interference resonator structure. In Mach Zehnder interferometer two light beams pass through the two arms of the MZ structure, refractive index in one of the arms is varied through some modulation mechanism, resulting in different optical path lengths. The superposition of the two beams at the other end will result in modulation of optical signal. [17]. Modulators using MZ structure are large in size and have slow transition time.

The use of resonator structure enables modulation using compact devices. The transmittivity of a resonator is given approximately by

$$T = \frac{T_{max}}{1 + \left(\frac{4nL}{\pi\Delta\lambda}\right)^2 \sin^2\left(\frac{2\pi nL}{\lambda}\right)}$$

where λ is the wavelength of light incident on the resonator, and $\Delta\lambda$ is the full-width at half maximum of the transmission at resonance. The resonator circulates light within the cavity at the resonance wavelength which increases the optical path length without increasing the physical device length. If the incident wavelength is resonance wavelength light exits the device with transmittivity $T = T_{max}$ after a photon life time. At all other wavelengths light destructively interferes in the cavity and is not transmitted through the resonator. From equation 1 it is observed that if the refractive index is changed the transmission changes, resulting in amplitude modulation of the signal transmitted through the resonator.

Absorptive modulators based on electroabsorption effect work by changing the optical absorption in the modulator structure by application of electric field which is obtained by reverse biasing PIN diodes that contain the bulk semiconductor or quantum well materials in the intrinsic region of the diodes. Many Quantum-well modulator devices have been demonstrated at high speeds employing only a few micrometers of optical path length even without the use of resonators [18-20], Although use of cavities can enhance the performance of these devices. Another approach is to bond III-V devices to silicon in a waveguide confi-

guration. Resulting in InAlGaAs QCSE modulators bonded to silicon structures[21].

Recently electroabsorption modulators employing Ge quantum well structures grown on silicon has been demonstrated[21]. A lot of research is being done on optimizing device structures and developing integration techniques for these Si and Ge based high speed and low power modulators [21-29] which are compatible with current CMOS technology.

2.3 Receiver

Receiver section in an optical interconnect comprises of a semiconductor Photo Detector followed by an electronic amplifier. The photo detector performs the operation of optical-to-electrical signal conversion. It is basically a reverse biased device which absorbs the incident

radiation and generates electron-hole pairs which in turn produces a photo current in the external circuit . The detector to be used in optical interconnects should provide high bandwidth, high sensitivity and easy optical coupling, moreover it should be amenable to high density fabrication. There are many semiconductor photo detectors that may fulfil these requirements: p-i-n photodiodes (PIN), metal-semiconductor-metal photodiodes (MSM), and avalanche photodiodes (APD). PINs are the most commonly used photo detector in short-distance optical links. MSM is a low-capacitance optical detector with high-speed operation and a larger active area than PIN but the responsivity of MSM is typically lower than of PIN and although APD provides the highest responsivity due to the internal gain, its drawbacks are the requirement of high-bias voltage and expensive fabrication process.

GaAs is a suitable material for photodiodes at wavelengths up to 850 nm and GaAs is more popular in high-speed applications but it is expensive. Si is used at wavelengths up to 1 μm ; In longer wavelengths up to 1.7 μm , an InGaAs material system on InP substrates is commonly used, but for optical interconnects the most important detector material is Ge, because of its potential to monolithically integrate photo diodes on Si-based integrated circuits. For optical interconnects waveguides are used as routing device thus, instead of being surface-normal, the Photo Diode needs to be waveguide integrated. Recently Waveguide-integrated Ge photo diodes have become a topic of intense research and development [30-41].

The small photocurrent generated by the Photo Diode must be amplified with a minimum amount of added noise for further processing. Hence, a preamplifier is

used as the first stage of amplification. Three widely used configurations for pre amplifier circuits are the high input impedance amplifier, the low input-impedance amplifier, and the trans impedance amplifier. TIA is typically chosen in optical interconnects because it meets the requirement of large bandwidth at low noise.

2.4 Wave guides, Switches, Routers

Optical waveguide to be used in optical interconnects must have a very low attenuation and good optical properties with high stability against stresses involved in the electronics assembly processes. The attenuation in silicon waveguides arises mainly due to light scattering from the etched sidewalls. Minimizing the optical field overlap with etched interfaces can effectively reduce the attenuation in waveguide. Increasing waveguide width and decreasing etch depth minimizes this overlap. Hence ridge waveguide structure is generally preferred for waveguides in optical interconnects. Several waveguide manufacturing technologies exist such as photolithography, printing, ion-exchange, laser direct writing, and laser ablation. The materials used for fabrication of waveguides include acrylates, polyimides, cyclic olefins, siloxanes and silsesquioxanes. [43-49]. Apart from waveguide other passive waveguide components, such as bends, splitters and crossings are essential for signal routing on the board.

3. CONCLUSIONS

The motivation for using optical interconnects technology is that the electrical interconnects cannot keep track of Moore's law indefinitely with increasing data rate requirements of high performance computers back frame computers , supercomputers and other high speed data processors . One possible alternative is use of optical interconnects. The main advantages offered by optical interconnects over their electrical counterparts are higher carrier frequency, less attenuation, less crosstalk and lower power consumption. Despite these advantages, there are a number of issues to be resolved before optical interconnect technology can be implemented.

Currently there are number of optical interconnect technologies at various stages of development. III-V quantum wells are considered the most mature technology which is already in the market. High index contrast structures are also being integrated in various chip configurations. The rest of the technologies are still in either fundamental research or applied research

phase and hence are not ready for the market. A lot of research and is still required for enhancing their performance. The incompatibility of mature III-V semiconductor materials with current CMOS technology is a major hurdle in economic production of reliable integrated circuits employing optical interconnects. Integration of components and manufacturing integrated circuits using optical interconnects requires significant research investments. The fact remains that silicon based technology which is compatible with current integrated circuit technology still need to be proven efficient technology for optical interconnects. The performance and reliability of available optical components needs to be improved significantly, moreover electrical interconnect technology is still good enough thus for most applications, optical interconnects are used for links requiring the highest data rates.

ACKNOWLEDGMENT

The first author Sumita Mishra is grateful to Maj. Gen. K.K. Ohri , Prof. S.T.H. Abidi and Brig. U. K. Chopra of Amity University, India for their support during the research work.

REFERENCES

1. M. J. Kobrinsky, B. A. Block, J.-F. Zheng, B. Barnett, E. Mohammed, M. Reshotko, F. Robertson, S. List, I. Young, and K. Cadien, "On-chip optical interconnects," *Intel Technol. J.*, vol. 8, no. 2, pp. 129-142, May 2004.
2. W. Green, S. Assefa, A. Rylyakov, C. Schow, F. Horst, Y. Vlasov, SEMICON 2010.
3. S. Assefa, W. Green, A. Rylyakov, C. Schow, F. Horst, Y. Vlasov, OMM6, OFC 2010.
4. J. Van Campenhout, et al., *Optics Letters* 35(7), 1013-1015 (2010).
5. J. Van Campenhout, et al., *Optics Express* 17(26), 24020-24029 (2009).
6. S. Assefa, et al., *IEEE Journal of Selected Topics in Quantum Electronics*, 16(5) 1376-1385 (2010).
7. S. Assefa, et al., *Optics Express* 18(5), 4986-4999 (2010).
8. V. R. Almeida, R. Panepucci, and M. Lipson, "Compact mode conversion for highly-confined waveguides," *Integrated Photonics Research (IPR) Conference*, paper IMD5-1, pp. 230-233, Washington 2003.
9. Koyama, F. "Recent Advances of VCSEL Photonics". *Journal of Lightwave Technology*, Vol. 24, No. 12, pp. 4502-4513. 2006.
10. D.A. Louderback, "Monolithically Integrated Vertical Cavity Lasers and Resonant Detectors for Free-Space Optical Interconnects," Ph.D. Dissertation, Department of Electrical and Computer Engineering, University of California, Santa Barbara, June 2003.
11. Cheng, J., Shieh, C.-L., Huang, X., Liu, G., Murty, M. V. R., Lin, C. C., Xu, D. X. 2005. Efficient CW lasing and high-speed modulation of 1.3 μm AlGaInAs VCSELs with good high temperature lasing performance. *IEEE Photonics Technology Letters*, Vol. 17, No. 1, pp. 7-9.
12. Thierry Pinguet, Vikram Sadagopan, Attila Mekis, Behnam Analui, Daniel Kucharski, Steffen Gloeckner "A 1550 nm, 10 Gbps optical modulator with integrated driver in 130 nm CMOS" *Proc 4th Conf Group IV Photon Tokyo Japan Sep (2007)*
13. R. A. Soref and B. R. Bennett, "Electrooptical effects in silicon," *IEEEJ. Quantum Electron.*, vol. 23, no. 1, pp. 123-129, Jan. 1987.
14. L. V. Keldysh, "The effect of a strong electric field on the optical properties of insulating crystals," *Soviet Phys.-JETP*, vol. 34, pp. 788-790, 1958
15. Y.-H. Kuo, Y. K. Lee, Y. Ge, S. Ren, J. E. Roth, T. I. Kamins, D. A. B. Miller, and J. S. Harris, "Strong quantum-confined stark effect in germanium quantum-well structures on silicon," *Nature*, vol. 437, pp. 1334-1336, Oct. 2005.
16. Juthika Basak, Ling Liao, Ansheng Liu, Doron Rubin, Yoel Chetrit, Hat Nguyen, Dean Samara-Rubio, Rami Cohen, Nahum Izhaky, and Mario Paniccia, "Developments in Gigascale Silicon Optical Modulators Using Free Carrier Dispersion Mechanisms" *Advances in Optical Technologies 2009*.
17. C. K. Tang and G. T. Reed, "Highly efficient optical phase modulator in SOI waveguides," *Electron. Lett.*, vol. 31, no. 6, pp. 451-452, Mar. 1995.
18. M. W. Geis, S. J. Spector, R. C. Williamson, and T. M. Lyszczarz, "Submicrosecond submilliwatt silicon-on-insulator thermo-optic switch," *IEEE Photon. Technol. Lett.*, vol. 16, no. 11, pp. 2514-2516, Nov. 2004
19. E. L. Wooten, K. M. Kissa, A. Yi-Yan, E. J. Murphy, D. A. Lafaw, P.F. Hallemeier, D. Maack, D. V. Attanasio, D. J. Fritz, G. J. McBrien, and D. E. Bossi, "A

review of lithium niobate modulators for fiber-optic communications systems," *IEEE J. Select. Topics. Quantum Electron.*, vo.6, no.1, pp. 69-82, Jan/Feb 2000.

20. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rublin, O. Cohen, R. Nicolaescu, and M. Paniccia, "A high-speed silicon optical modulator based on a metal-oxide-semiconductor capacitor," *Nature*, vol. 427, pp. 615-618, Feb. 2004

21. Q. Xu, B. Schmidt, S. Pradhan, and M. Lipson, "Micrometre-scale silicon electro-optic modulator," *Nature*, vol. 435, pp. 325-327, May 2005.

22. Liu, L. Liao, D. Rubin, J. Basak, Y. Chetrit, H. Nguyen, R. Cohen, N. Izhaky, and M. Paniccia, B "Recent development in a high-speed silicon optical modulator based on reverse-biased pn diode in a silicon waveguide" *Semicond. Sci. Technol.*, vol. 23, p. 064001, 2008

23. M. B. Yairi, C. W. Coldren, D. A. B. Miller, and J. S. Harris, Jr., "High-speed, optically controlled surface-normal modulator based on diffusive conduction," *Appl. Phys. Lett.* vol. 75, no. 5, pp. 597-599, 1999.

24. M. Chacinski, U. Westergren, B. Willen, J. Stoltz, and L. Thylen, "Electroabsorption modulators suitable for 100 Gb/s ethernet," *IEEE Electron Device Lett.*, vol. 29, pp. 1014-1016, 2008.

25. S. Kodama, T. Yoshimatsu, and H. Ito, "500 Gbit/s optical gate monolithically integrating photodiode and electroabsorption modulator," *Electron. Lett.*, vol. 40, pp. 555-556, 2004.

26. Jifeng Liu, Mark Beals, Andrew Pomerene, Sarah Bernardis, Rong Sun, Jing Cheng, Lionel C. Kimerling and Jurgen Michel, "Waveguide-integrated, ultralow-energy GeSi electro-absorption modulators" *Nature Photonics*, Vol 2 pp433-437, Jul, 2008.

27. J. Di Liang, "Recent progress in lasers on silicon," *Nature Photon.*, vol.4, no. 8, pp. 511-517, 2010.

28. G. Reed, G. Mashanovich, F. Gardes, and D. Thomson, "Silicon optical modulators," *Nature Photon.*, vol. 4, no. 8, pp. 518-526, 2010.

29. L. Liao, D. Samara-Rubio, M. Morse, A. Liu, D. Hodge, D. Rubin, U. D. Keil, and T. Franck, "High speed silicon Mach-Zehnder modulator," *Opt. Express*, vol. 13, no. 8, pp. 3129-3135, Apr. 2005.

30. R. Soref, "Mid-infrared photonics in silicon and germanium," *Nature Photon.*, vol. 4, no. 8, pp. 495-497, 2010.

31. Konstantin Nikolic, Alexandru Serb, and Timothy G. Constandinou, "An Optical Modulator in Unmodified, Commercially Available CMOS Technology" *IEEE photonics technology letters*, vol. 23, no. 16, August 2011

32. M. Beals, J. Michel, J. F. Liu, D. H. Ahn, D. Sparacin, R. Sun, C. Y. Hong, L. C. Kimerling, A. Pomerene, D. Carothers, J. Beattie, A. Kopa, A. Apsel, M. S. Rasras, D. M. Gill, S. S. Patel, K. Y. Tu, Y. K. Chen, and A. E. White, "Process flow innovations for photonic device integration in CMOS," *SPIE Proc.*, vol. 6898, pp. 1-14, Feb. 2008.

33. D. Ahn, C. Hong, J. Liu, W. Giziewicz, M. Beals, L. C. Kimerling, and J. Michel, "High performance, waveguide integrated Ge photodetectors," *Opt. Exp.*, vol. 15, pp. 3916-3921, Mar. 2007.

34. T. Yin, R. Cohen, M. M. Morse, G. Sarid, Y. Chetrit, D. Rubin, and M. J. Paniccia, "31GHz Ge n-i-p waveguide photodetectors on silicon-on-insulator substrate," *Opt. Exp.*, vol. 15, pp. 13965-13971, Oct. 2007.

35. J. Wang, W. Y. Loh, K. T. Chua, H. Zang, Y. Z. Xiong, S. M. F. Tan, M. B. Yu, S. J. Lee, G. Q. Lo, and D. L. Kwong, "Low-voltage high-speed (18GHz/1V) evanescent-coupled thin-film-Ge lateral PIN photodetectors integrated on Si waveguide," *IEEE Photon. Technol Lett.*, vol. 20, no. 17, pp. 1485-1487, Sep. 2008.

36. L. Vivien, J. Osmond, J. F'ed'eli, D. Marris-Morini, P. Crozat, J. Damlencourt, E. Cassan, Y. Lecunff, and S. Laval, "42 GHz p-i-n germanium photodetector integrated in silicon-on-insulator waveguide," *Opt.mExp.*, vol. 17, pp. 6252-6257, Apr. 2009.

37. J. F. Liu, D. Pan, S. Jongthammanurak, D. Ahn, C. Y. Hong, M. Beals, L. C. Kimerling, J. Michel, A. T. Pomerene, C. Hill, M. Jaso, K. Y. Tu, Y. K. Chen, S. Patel, M. Rasras, A. White, and D. M. Gill, "Waveguide- Integrated Ge p-i-n photodetectors on SOI Platform," in *Proc. 3rd IEEE Int. Conf. Group IV Photon.*, May 2006, pp. 173-175.

38. C. Gunn, "CMOS photonics for high-speed interconnects," *IEEE Micro*, vol. 26, no. 2, pp. 58-66, Mar./Apr. 2006.

39. G. Masini, S. Sahni, G. Capellini, J. Witzens, and C. Gunn, "Waveguide photodetectors integrated in a CMOS process," *Adv. Opt. Technol.*, pp. 196572-1-196572-5, 2008.

40. S. Assefa, F. Xia, S. W. Bedell, Y. Zhang, T. Topuria, P. M. Rice, and Y. A. Vlasov, "CMOS-integrated 40GHz germanium waveguide photodetector for on-chip optical interconnects," presented at the *Opt. Fiber Commun. Conf. (OFC)*, San Diego, CA, Paper OMR4, Feb. 2009.

41. S. Assefa, F. Xia, S. W. Bedell, Y. Zhang, T. Topuria, P. M. Rice, and Y. A. Vlasov, "CMOS-integrated germanium waveguide photodetector with ultralow capacitance for 40Gbps transceivers," *Opt. Exp.*, vol. 18, pp. 4986-4999, Mar. 2010.
42. Kromer, C., Sialm, G., Berger, C., Morf, T., Schmatz, M.L., Ellinger, F., Erni, D., Bona, G.-L., Jackel, H. 2005.
43. Y. A. Vlasov, and S. J. McNab, "Losses in single-mode silicon-on-insulator strip waveguides and bends," *Opt. Express* 12, 1622-1631 2004.
44. M. Gnan, S. Thoms, D. S. Macintyre, R. M. De La Rue, and M. Sorel, "Fabrication of low-loss photonic wires in silicon-on-insulator using hydrogen silsesquioxane electron-beam resist," *Electron. Lett.* 44, 115-116 2008
45. M. A. Webster, R. M. Pafchek, G. Sukumaran, and T. L. Koch, "Low-loss quasi-planar ridge waveguides formed on thin silicon-on-insulator," *Appl. Phys. Lett.* 87, 2005.
46. L. K. Rowe, M. Elsey, N. G. Tarr, A. P. Knights, and E. Post, "CMOS-compatible optical rib waveguides defined by local oxidation of silicon," *Electron. Lett.* 43, 392-393 , 2007.
47. F. Y. Gardes, G. T. Reed, A. P. Knights, G. Mashanovich, P. E. Jessop, L. Rowe, S. McFaul, D. Bruce, and N. G. Tarr, "Sub-micron optical waveguides for silicon photonics formed via the Local Oxidation of Silicon (LOCOS), *Proc. of SPIE* 6898, 68980R-68980R-4 , 2008.
48. R. Pafchek, R. Tummidi, J. Li, M. A. Webster, E. Chen, and T. L. Koch, "Low-loss silicon-on-insulator shallow-ridge TE and TM waveguides formed using thermal oxidation," *Appl. Opt.* 48, 958-963 , 2009.
49. J. Cardenas, C. B. Poitras, J. T. Robinson, K. Preston, L. Chen, and M. Lipson, "Low loss etchless silicon photonic waveguides," *Opt. Express* 17, 4752-4757 , 2009.

COMPARISON ANALYSIS OF DIFFERENT CONTROLLERS FOR PWM INVERTER FED PERMANENT MAGNET BRUSHLESS DC MOTOR

P.Elangovan, Dr.C.Kumar

Abstract — This paper presents the performance evaluation of the fuzzy and Proportional and Integral control system applied to Permanent-Magnet Brushless DC Motor (PMBLDCM) with different Pulse Width Modulation(PWM) techniques. At first the comparison analysis is made between Sinusoidal PWM and Space Vector PWM inverter fed Permanent-Magnet Brushless DC Motor with the implementation of Proportional and Integral controller. From the analysis of Total harmonic distortion (THD) in the inverter current, Space Vector Pulse Width Modulation (SVPWM) technique was found to be better than Sinusoidal Pulse Width Modulation(SPWM) technique. However, the PI controller has some disadvantages such as: high starting overshoot, sensitivity to controller gains and sluggish response due to sudden load disturbance. Hence the Fuzzy logic controller is implemented in the feedback for SVPWM inverter fed PMBLDCM. Simulation result are presented and analyzed for both fuzzy and PI controllers. It is observed that fuzzy logic based controller gives better responses than traditional Proportional and Integral controller for the speed control of dc motor drives.

IndexTerms — Pulse Width Modulation ; Permanent Magnet Brushless DC Motor ; PI controller; Fuzzy logic Controller; Sinusoidal Pulse Width Modulation ; Space Vector Pulse Width Modulation ; Total Harmonic Distortion.

1 INTRODUCTION

Conventional dc motors are highly efficient and their characteristics make them suitable as servomotor. However, it needs a commutator and brushes which are subject to wear and required maintenance. The functions of commutator and brushes were implemented by solid-state switches that can realize maintenance-free motors. These motors are now known as brushless dc motors. Brushless dc motors are widely used in various applications. Two examples of them are electric vehicle and industrial machinery. Fuzzy logic controller which is presented by Zadeh in 1965, is a new controller [1]. Besides that, fuzzy logic controller is more efficient from the other controller such as proportional-integral (PI) controller. The comparison between them is needed to compare in what way the controller is efficient [2]. The reason why conventional controller is having low efficiency such as PI controller because the overshoot is too high from the set point and it may takes delay time to get constant and sluggish response due to sudden change in load torque and the sensitivity to controller gains K_i and K_p [3].

PWM techniques:

Pulse Width Modulation variable speed drives are increasingly applied in many new industrial applications that require superior performance. Recent developments in power electronics and semiconductor technology have lead improvements in power electronic systems. Hence, different circuit configurations namely multilevel inverters have become popular and considerable interest by researcher are given on them. Variable voltage and frequency supply to d.c drives is invariably obtained from a three-phase voltage source inverter. A number of Pulse width modulation (PWM) schemes are used to obtain variable voltage and frequency supply. The most widely used PWM schemes for three-phase voltage source inverters are carrier-based sinusoidal PWM and space vector PWM. There is an increasing trend of using space vector PWM because of their easier digital realization and better dc bus utilization.

2 PULSE WIDTH MODULATION IN INVERTERS

Output voltage from an inverter can also be adjusted by exercising a control within the inverter itself. The most efficient method of doing this is by pulse-width modulation control used within an inverter. In this method, a fixed dc input voltage is given to the inverter and a controlled ac output voltage is obtained by adjusting the on and off

• P.Elangovan, M.E., Asst.Prof/ EEE, SKP Engineering College, Tiruvannamalai, TamilNadu, India. E-mail: elangoeee2007@gmail.com

• Dr.C.Kumar,B.E.,M.sc.,(Engg),P.hd., Director (Academic), SKP Engineering College, Tiruvannamalai, TamilNadu, India. E-mail: drchkumararima@gmail.com

periods of the inverter components. This is the most popular method of controlling the output voltage and this method is termed as Pulse-Width Modulation Control.

PWM inverters are quite popular in industrial applications. PWM techniques are characterized by constant amplitude pulses. The width of these pulses is however modulated to obtain inverter output voltage control and to reduce its harmonic content. The different PWM techniques are as under:

- (a) Single-pulse modulation
- (b) Multiple pulse modulation
- (c) Sinusoidal pulse width modulation

The carrier based PWM technique has been discussed in [1] and the corresponding equations related to it has been derived.

2.1 Sinusoidal PWM technique

The generation of gating signals with sinusoidal PWM are shown in fig.1. There are three sinusoidal reference signal corresponding to a phase to generate the gating signals for that phase [8]. Comparing the carrier signal with the reference phases V_{oa} , V_{ob} and V_{oc} produces the gating signals. The instantaneous line-to-line output voltage is $V_{ab} = V_s(g_1 - g_2)$. The output voltage as shown in fig.1 is generated by eliminating the condition that two switching devices in the same arm cannot conduct at the same time.

The normalized carrier frequency m_r should be odd multiple of three. Thus, all phase-voltage are identical, but 120° out of phase without even harmonics; moreover, harmonics at frequencies multiple of three are identical in amplitude and phase in all phases. For instance, if the ninth harmonic voltage in phase a is

$$V_{a9}(t) = \tilde{v}_9 \sin(9\omega t) \dots \dots \dots (1)$$

the corresponding ninth harmonic in phase b will be,

$$V_{b9}(t) = \tilde{v}_9 \sin(9(\omega t - 120^\circ)) \\ = \tilde{v}_9 \sin(9(\omega t - 1080^\circ)) = \tilde{v}_9 \sin(9\omega t) \dots \dots (2)$$

Thus, the ac output line voltage V_{ab} does not contain the ninth harmonic. Therefore, for odd multiples of three times the normalized carrier frequency m_r , the harmonics in the ac output voltage appear at normalized frequencies f_h centered around m_r and its multiples, specifically, at

$$N = jm_r \pm k \dots \dots \dots (3)$$

For nearly sinusoidal ac load current, the harmonics in the dc link current are at frequencies given by

$$n = jm_r \pm k \pm 1 \dots \dots \dots (4)$$

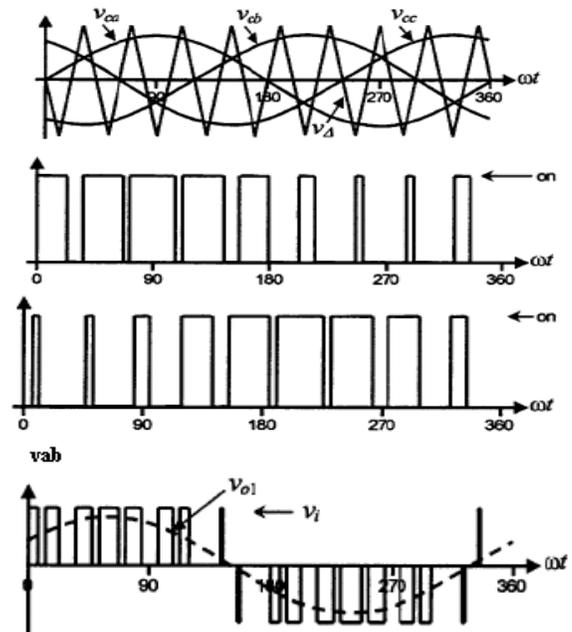


Fig.1 SPWM for three phase inverter

2.2 Space Vector PWM technique

The desired three phase voltages at the output of the inverter could be represented by an equivalent vector \mathbf{V} rotating in the counter clock wise direction as shown in Fig.2. The magnitude of this vector is related to the magnitude of the output voltage and the time of this vector takes to complete one revolution is the same as the fundamental time period of the output voltage.

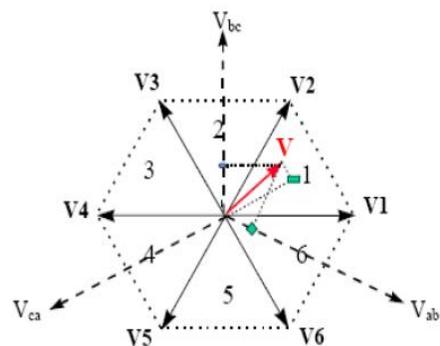


Fig.2 Output voltage vector in the plane.

Let us consider the situation when the desired line-to-line output voltage vector \mathbf{V} is in sector 1 as shown in Fig.3.

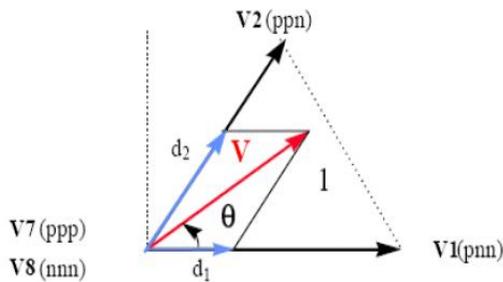


Fig.3 Synthesis of the required output voltage vector in sector 1.

This vector could be synthesized by the pulse-width modulation (PWM) of the two adjacent SSV's **V1** (pnn) and **V2** (ppn), the duty cycle of each being d_1 and d_2 , respectively, and the zero vector(**V7**(nnn) / **V8**(ppp)) of duty cycle d_0 :

$$d_1V_1 + d_2V_2 = V = mV_{ge} \dots\dots\dots (5)$$

$$d_1 + d_2 + d_0 = 1 \dots\dots\dots (6)$$

Where, $0 \leq m \leq 0.866$, is the modulation index. This would correspond to a maximum line-to-line voltage of $1.0V_g$, which is 15% more than conventional sinusoidal PWM as shown.

All SVM schemes and most of the other PWM algorithms use Eqns. (5) and (6) for the output voltage synthesis. The modulation algorithms that use non-adjacent SSV's have been shown to produce higher THD and/or switching losses and are not analyzed here, although some of them, e.g. hysteresis, can be very simple to implement and can provide faster transient response. The duty cycles d_1 , d_2 , and d_0 , are uniquely determined from Eqns. (5) and (6), the only difference between PWM schemes that use adjacent vectors is the choice of the zero vector(s) and the sequence in which the vectors are applied within the switching cycle.

3 STRUCTURES

3.1 Proportional-Integral (PI) Controller Structure

Fig. 4 shows the Proportional-Integral (PI) controller block diagram. The speed error E_N between the reference speed N_R and the actual speed N of the motor is fed to the PI controller, and the K_1 and K_2 are the proportional and integral gains of the PI controller.

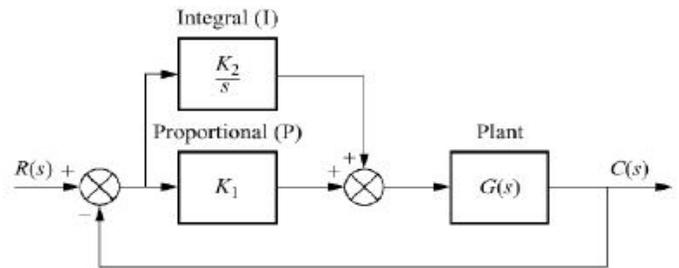


Fig.4 Block Diagram of PI controller

The P-I controller has the form

$$E_1(S)/E_N(S) = (K_1S+K_2)/S \dots\dots (7)$$

This is a phase-lag type of controller with the pole at the origin and makes the steady-state error in speed zero. The transfer function between the output speed N and the reference speed N_R is given by:

$$N(S)/N_R(S) = (AK_1+AK_2S)/(K_1S^2+K_2S+K_3) \dots\dots(8)$$

Where, $A = C_1K_{CH}K$

$$K_1 = R_A B T_M + C_1 K_{CH} B T_M$$

$$K_2 = R_A B + K_2 + C_1 K_{CH} B + A K_P$$

$$K_3 = A K_I$$

$$T_M = J / B$$

K_I and K_P are controller gains, and R_A , B , T_M , etc., are motor and feedback constants shown in table 1. The above equation introduces a zero and therefore a higher overshoot is expected for a step change in speed reference.

3.2 Fuzzy logic controller structure

Fig. 5 shows the basic structure of fuzzy logic controller. Fuzzy logic's linguistic terms are most often expressed in the form of logical implications, such as If-Then rules. These rules define a range of values known as fuzzy membership functions [2]. Fuzzy membership functions may be in the form of triangle, a trapezoid, a bell or another appropriate form.

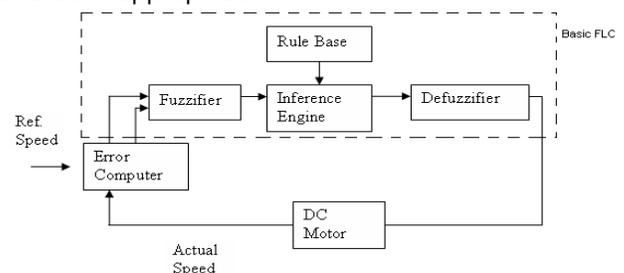


Fig. 5 Fuzzy logic controller

The inputs of the fuzzy controller are expressed in several linguistic levels shown in fig.6, these levels can be described as positive big (PB), positive medium (PM), positive small

(PS), or in other levels. Each level is described by a fuzzy set.

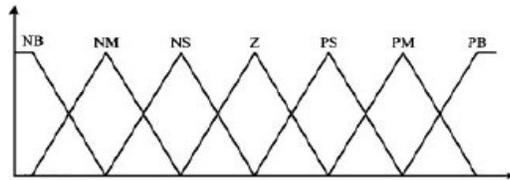


Fig.6 Seven levels of fuzzy membership function

In general, experience and expertise are required for the implementation of fuzzification in complex systems [3].

Fuzzy logic control doesn't need any difficult mathematical calculation, it only use simple mathematical calculation, but it can provide very good performance in a control system [4]. Thus, it can be one of the best available answers today for a board class of challenging controls problem. A fuzzy logic control consists of

- (i) Fuzzification: This process converts or transforms the measured inputs called crisp values, into the fuzzy linguistic values used by the fuzzy reasoning mechanism.
- (ii) Knowledge Base: A collection of the expert control rules (knowledge) needed to achieve the control goal.
- (iii) Fuzzy Reasoning Mechanism: This process will perform fuzzy logic operations and result the control action according to the fuzzy inputs.
- (iv) Defuzzification unit: This process converts the result of fuzzy reasoning mechanism into the required crisp value.

4 SIMULATION

To validate the control strategies as described, digital simulation were carried out on a converter dc motor drive system by using MATLAB/SIMULINK, the used parameters in these system are given in table 1.

Armature resistance (Ra)	0.5 Ω
Armature inductance (La)	8 mH
Back e.m.f constant (K)	0.55 V/rad/s
Mechanical inertia (J)	0.0465 kg.m ²
Friction coefficient (B)	0.004 N.m/rad/s
Rated armature current (Ia)	10 A

Table 1. The parameter of dc motor drive system

4.1 Simulation results for SPWM and SVPWM techniques:

The THD when SPWM technique used is 20% whereas the THD when SVPWM technique used is only 2%. On observing fig.7 and fig.8, it is clear that the harmonic contents in the inverter current is more eliminated using SVPWM technique than SPWM technique.

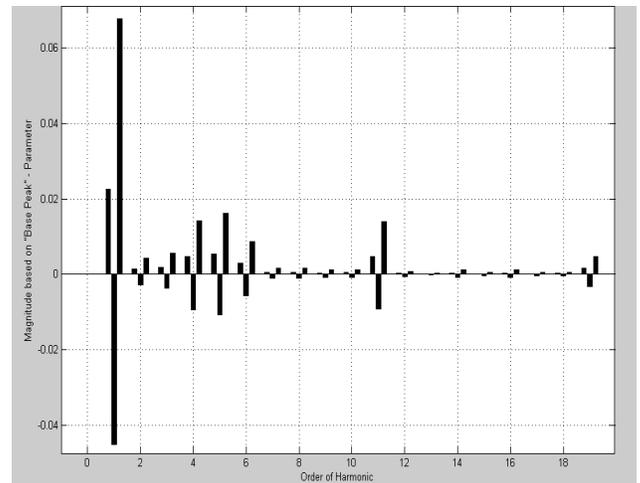


Fig.7 THD in current when SPWM technique used

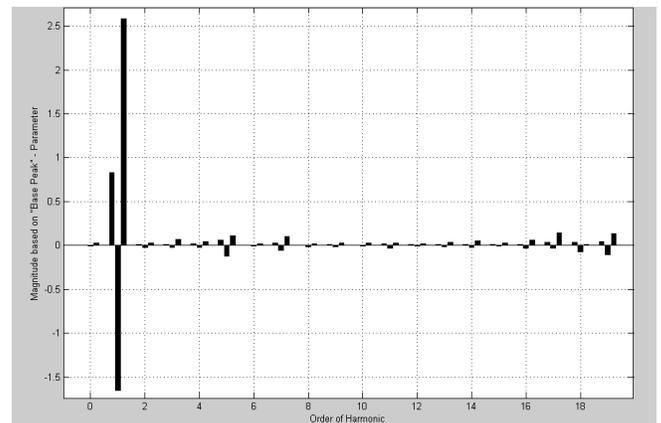


Fig.8 THD in current when SVPWM technique used

4.2 Simulation results for PI and Fuzzy logic controller:

The response of the drive system is obtained by setting the reference speed to 1500 r.p.m. The system speed response of fuzzy logic controller is shown in fig. 9. Fig. 11 shows the system speed response with fuzzy controller due to the load change and the load reference is 1400r.p.m. Compare to PI controller, it seen no starting overshoot and sudden load change problem for fuzzy logic controller. The system speed response of PI controller is shown in fig. 10. Fig. 12 shows the system speed response with PI controller due to the load change from 1500 r.p.m to 1400 r.p.m. Clearly observe that, the speed response of PI controller shows in fig. 10 has high starting overshoot from the set point. It approximate to 1900 r.p.m.

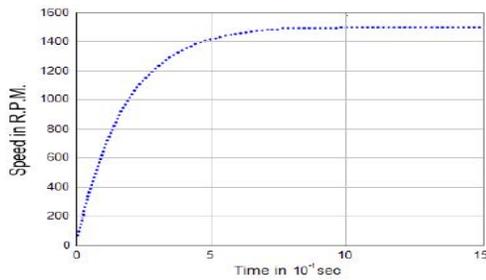


Fig.9 Speed response of fuzzy logic controller ($N_{ref} = 1500$ r.p.m)

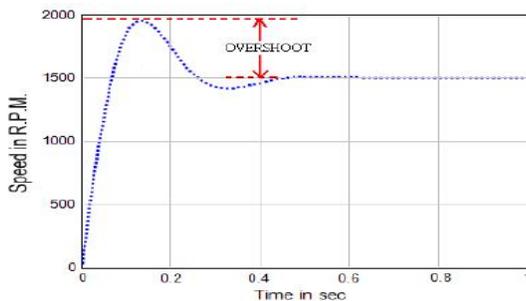


Fig.10 Speed response of PI controller ($N_{ref} = 1500$ r.p.m)

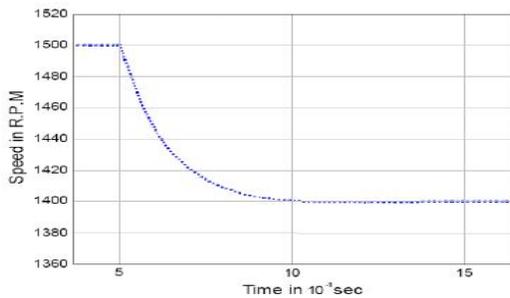


Fig.11 Speed response of fuzzy logic controller ($N_{ref} = 1400$ r.p.m)

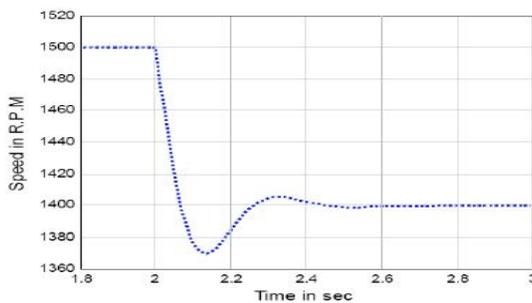


Fig.12 Speed response of PI controller ($N_{ref} = 1400$ r.p.m)

5 CONCLUSION

This paper is intended to evaluate the performance of different controllers (PI and Fuzzy logic controller) when they are used in the feedback path of PWM inverter fed PMBLDC drive. In this paper, the performance evaluation of SPWM and SVPWM techniques also

determined based on the Total harmonic distortion in the inverter output current. From the simulation results, it is concluded that SVPWM technique is more preferable than SPWM technique to control the output voltage and current of the inverter.

It is observed that fuzzy logic controller provide important advantages over the traditional PI controller like limiting the overshoot in speed, thus the starting current overshoot can be reduced. This paper also demonstrates the successful application of fuzzy logic control to a phase controlled converter dc motor drive. Fuzzy logic was used in the design of speed controllers of the drive system and the performance was compared with that of PI controller.

The advantages of the Fuzzy controller are that it determines the number of rules automatically, reduces computational time, learns faster and produces lower errors than other method. By proper design a fuzzy logic controllers is much better than PI controllers for the speed control of dc motor drives.

REFERENCES

- [1] "Control of Voltage Source Inverters using PWM/SVPWM for Adjustable Speed Drive Applications" Prof. B. Chitti Babu, NIT, Rourkela, May 2009
- [2] Hanns, Michael. "A Nearly Strict Fuzzy Arithmetic for Solving Problems with Uncertainties." Online posting. 14 Dec. 2002.
- [3] "Tour Of Fuzzy Logic Functions." Wolfram Reasearch, Inc. 28 Nov.2002.
- [4] Espinosa J.J. and Vandewalle J., "Predictive Control Using Fuzzy Models", Advances in Soft Computing - Engineering Design and Manufacturing (Eds.), R. Roy, Springer - London, ISBN 1-852-33062-7, 1998
- [5] M.Ali Akcayol, Aydin Cetin, and Cetin Elmas, (November 2002), An Educational Tool for Fuzzy Logic-Controlled BDCM, IEEE Transaction on Education, vol. 45, no. 1.
- [6] Devendra Rai, Brushless DC Motor Simulink Simulator, Department of Electronics and Communication Engineering, National Institute of Technology Karnataka, Surathkal 575 025, INDIA
- [7] Yodyium Tipsuwan and Mo-Yuen Chow, "Fuzzy Logic microcontroller implementation for DC motor speed control", IEEE Trans. Power Electronics, Vol. 11, No.3, pp 1271-1276, 1999.
- [8] J.R.Espinoza, Power Electronics Handbook, edited by M.H.Rashid, San Diego, CA: Academic Press, 2001, chapter 14 - Inverters.
- [9] M. Depenbrock, "Pulsewidth control of a 3-phase inverter with nonsinusoidal phase voltages," in Proc. IEEE-IAS Int. Semiconductor Power Conversion Conf., Orlando, FL, 1975, pp. 389-398.
- [10] J. A. Houldsworth and D. A. Grant, "The use of harmonic distortion to increase the output voltage of a three-phase PWM inverter," IEEE Trans. Ind. Applicat., vol. 20, pp. 1224- 1228, Sept./Oct.1984.

Modulation Error as QoS Trigger for Optimum Vertical Handover : Analysis for Adhoc Network

Reena Sharma, Kuldeep Singh

ABSTRACT

Next generation network is envisaged to be a heterogeneous network with integration of different radio access network (RAN) technology (4G) i.e circuit switched and IP centric. To maintain a better Quality of Service (QoS) during a seamless transfer of a session from one RAN to another is termed as Vertical Handover. One of the key wireless network is Adhoc network which is a decentralized wireless network. This paper proposes a Measurement report comprising of minimum signal power (E_b/N_0) for different frames at different sublayers of PHY and MAC Layer. Here Error Vector Measurement (EVM rms %) at the receiver point is obtained which is exposed to hostile environment i.e impaired with AWGN channel with frequency offset and non-linear propagation (George) model. Measurement suggests De/authentication and RTS frames needs to be transmitted at higher SNR for achieving better QoS (i.e Zero FCS-error).

KEYWORD

AWGN,BER,CTS,EDGE,EVM,FCS,GSM,MAC,OFDM,RTS,SSID,SNR

1. INTRODUCTION

An ad hoc wireless network is a collection of wireless nodes that self-configure to form a network without the aid of any established infrastructure. Some or possibly all of these nodes are mobile.[2]

In this paper, Adhoc network is simulated on WiLANTA software which strictly adheres to IEEE802.11 b/g norms. Here the figure of merit for the modulation accuracy is the Error Vector Magnitude (EVM), which represents the distance between the measured and the perfect modulated signals[Fig.1]. EVM is used instead of the typical figure of merit, bit-error-rate (BER), because BER suffers from some limiting factors, such as the requirement for dedicated equipment, long measurement intervals (QoS) metrics on which network selection and connection is based and a limited diagnostic value. BER, Delay, Bandwidth etc are treated as Quality of Service metrics. Here EVM is used as a proposed QoS trigger to initiate a Handover. Therefore EVMrms is observed for different frames of Adhoc network and a comparative study of different frames/layers is done.

2. AD HOC WIRELESS LOCAL AREA NETWORK STANDARDS

IEEE Std 802.11g.-2003 was introduced in 2003.

Modulation format: It uses OFDM, CCK(Complimentary Code Keying), and BCC (Packet Binary Convolution Coding) modulation schemes.

Max data rate : 54 Mbps.

Operating frequency : 2.4 GHz .

Max power output : 1000 mw.

Compatibility : compatible with 802.11b .

No of Channels : 14 (3 non overlapping)

3. AD HOC NETWORK,MAC FRAME TYPE AND SUBFRAME TYPE

3.1 Adhoc Network : An Adhoc wireless network is a collection of wireless nodes that self-configure to form a network without the aid of any established infrastructure. Some or possibly all of these nodes are mobile. These networks are extremely compelling for applications where a communications infrastructure is too expensive to deploy, cannot be deployed quickly, or is simply not feasible. There are numerous potential applications for ad hoc wireless networks, ranging from multihop wireless broadband Internet access, to sensor networks, to building or highway automation, to voice, image, and video communication for disaster areas.[2]

-
- Assistant Professor , EEE Deptt. , Galgotias College of Engineering and Technology , Greater Noida ,U.P.,India
 - Assistant Professor , ECE Deptt. , Galgotias College of Engineering and Technology , Greater Noida ,U.P.,India

3.2 Management Frames : Management Frames contain information for the receiving MAC management entity. Management Frame type has the following Sub frame types.[1]

3.2.1. Probe request Frame : A station sends a probe request frame when it needs to obtain information

from another station. For example, a radio NICs/station would send a probe request to determine which access points are within range.

3.2.2. Probe response Frame : A station will respond with a probe response frame, containing capability information, supported data rates, etc., when after it receives a probe request frame.

3.2.3. Beacon Frame : The access point periodically sends a beacon frame to announce its presence and relay information, such as timestamp, SSID, and other parameters regarding the access point to radio NICs/stations that are within range. Radio NICs/stations continually scan all 802.11 radio channels and listen to beacons as the basis for choosing which access point is best to associate with.

3.2.4. Authentication Frame : A station sends an authentication frame to another station if it wishes to begin secure communication.

3.2.5. Deauthentication Frame : A station sends a deauthentication frame to another station if it wishes to terminate secure communication.

3.3 Control frames : Control frames contain information to control access to the wireless medium and assist in the delivery of data frames. Control frames type has the following Sub frame types.[1]

3.3.1. Power Save (PS)-Poll : When a station wakes up from a power save mode it transmits a Power Save Poll Frame to the access point to retrieve any frames buffered while it was in power save mode.

3.3.2. Request to Send (RTS) : A station sends a RTS frame to another station as the first phase of a two-way handshake necessary before sending a data frame.

3.3.3. Clear to Send (CTS) : A station responds to a RTS with a CTS frame, providing clearance for the requesting station to send a data frame.

3.4 Data frame : Data frames contain data from higher protocol layers as indicated by their name, but not always.

They can contain only data, data with control information; contain only PCF control information or sometimes no data at all.[1]

3.4.1. Data : Frames of the data subtype are used for the purpose of just transmitting the frame body or data from 802.11bg transmitter to 802.11bg receiver.

4. RF FRONT-END NON IDEALITIES or SIGNAL IMPAIRMENTS

4.1. AWGN (Additive White Gaussian Noise) : It is a common wideband channel thermal noise impairment, on which SNR (Signal to Noise Ratio) is typically based. If SNR is high, we can decode the transmitted signal easily. On the other hand, If SNR is low, decoded the noisy signal becomes difficult and prone to errors.

4.2. Frequency Offset : Frequency offset is the difference between the frequency of a source and a reference frequency/carrier frequency. The frequency offset occurs due to a mismatch of oscillator frequencies or Doppler shift which results from a relative movement between transmitter and receiver in a mobile environment. The frequency offset must be specified in Hz in the range (-125 KHz to 125 KHz) according to IEEE 802.11b/g standard.

4.3. Memoryless Nonlinearity : Memoryless Nonlinearity block applies a memoryless nonlinearity to a complex, baseband signal. It models radio frequency (RF) impairments to a signal at the receiver. George Model is used for modelling the nonlinearity.

5. ERROR VECTOR MAGNITUDE

3GPP standards provide the following definition of EVM: "The Error Vector Magnitude is a measure of the difference between the reference waveform and the measured waveform as illustrated by Figure 1. This difference is called the error vector[5]. EVM gives the measure of modulation error. The modulation error indicates the deviation of In phase and Quadrature phase (I/Q) values from ideal signal states and thus provides a measure of signal quality[1].

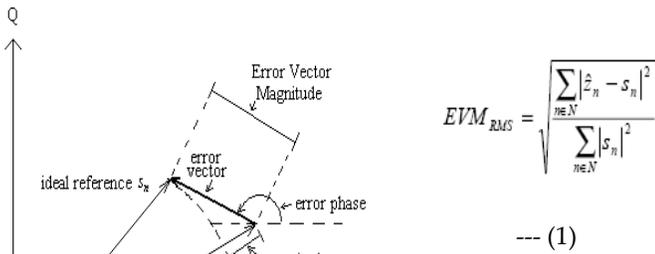


Figure 1: Error Vector Magnitude and related quantities.

The instantaneous error vector is subtracting the ideal reference from the modified version of the measured waveform. The root mean square EVM is obtained by above eq (1).

IEEE 802.11g Standard EVM Limits

$$Error_{RMS} = \frac{\sum_{i=1}^{N_f} \sqrt{\frac{\sum_{j=1}^{L_p} \left[\sum_{k=1}^{52} \{ (I(i,j,k) - I_0(i,j,k))^2 + (Q(i,j,k) - Q_0(i,j,k))^2 \} \right]}{52 L_p \times P_0}}}{N_f} \quad \text{--- (2)}$$

- L_p: Length of the packet
- N_f: Number of frames for the measurement
- I₀(I,j,k), Q₀(I,j,k) : ideal symbol point of ith frame, jth OFDM symbol of the frame, kth subcarrier of the OFDM symbol in the complex plane
- I(I,j,k), Q(I,j,k) : observed point of the ith frame, jth OFDM symbol of the frame, kth subcarrier of the OFDM symbol in the complex plane
- P₀: average power of the constellation. [10]

6. SIMULATION CONFIGURATION

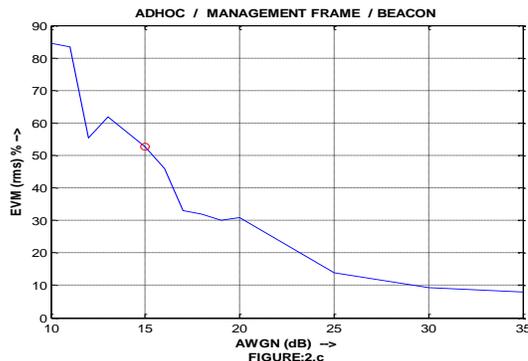
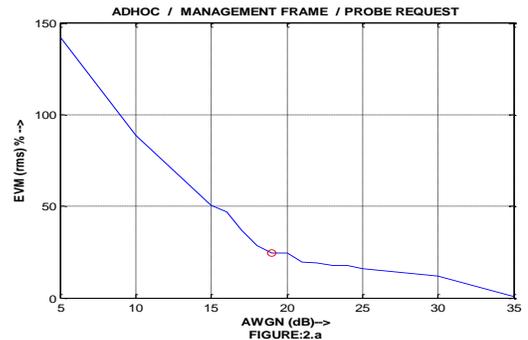
The simulation is done using the WiLANTA IQ generator and analyser software. The different parameters selected are as follow.

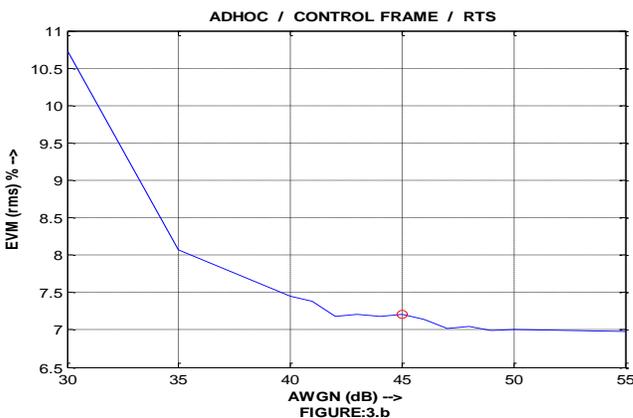
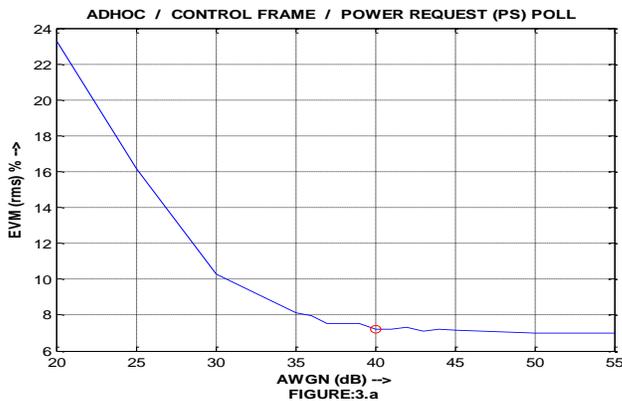
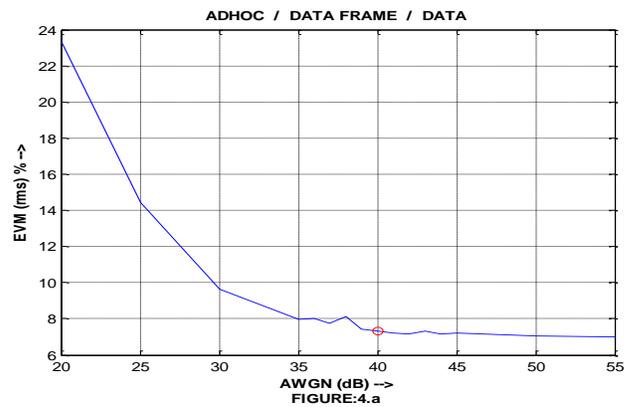
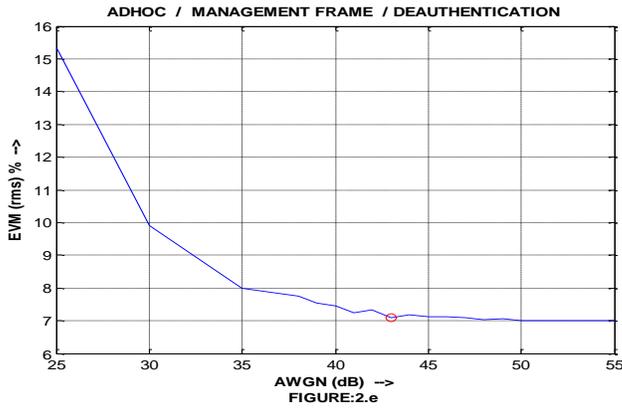
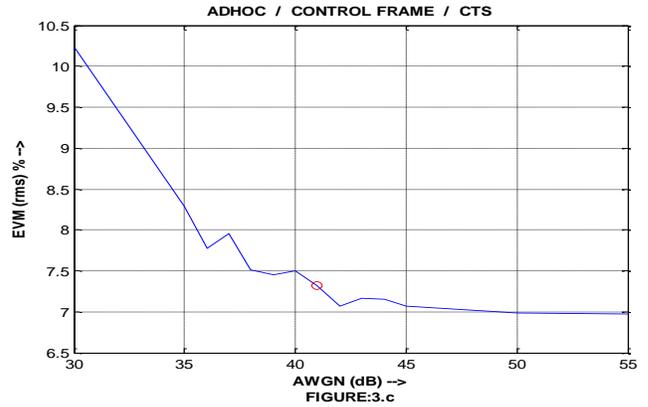
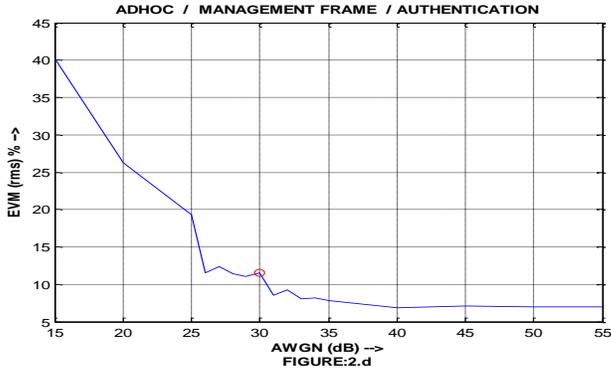
Parameter selected:

- Software : Wilanta IQ Generator IEEE
- 802.11 g;
- Standard : IEEE 802.11g
- Modulation scheme : OFDM
- Data Rate : 9 Mbps

- Scrambler : on
- Data Payload : 10Kbps
- Symbol/frame : 285
- Pre designed pattern : 00001111
- Packet count : 1
- MAC Parameters :
- Network type : Adhoc
- Frame type : Management / Control / Data
- More fragment : not required
- Power management : Active mode
- WEP Encryption : WEP not enabled
- Duration : 20 ms
- Impairments : AWGN
- Frequency offset : 10 KHz
- Memoryless Nonlinearity: George model

7. SIMULATION RESULT AND DESCRIPTION





1. An attempt has been made to obtain EVM rms, by varying AWGN. Management, Control and Data frames are used for the analysis as user has to associate with a new Access point for VHO decision. It can be observed that we get low EVMs when AWGN(dB) are slightly increased. Here we are testing our proposed metric in fading environment.

2. Eb/No (AWGN) dB for successful transmission (Zero FCS error) for different frames :

Network / Frame / Subtype	Min. Eb/No (dB)	EVMrms%
Adhoc / Management / Probe request	19	24.6124
Adhoc / Management / Probe response	20	22.6645
Adhoc / Management / Beacon	15	52.6606
Adhoc / Management / Authentication	30	11.4837
Adhoc / Management / Deauthentication	43	7.07913
Network / Frame /	Min.	EVMrms%

Subtype	Eb/No (dB)	
Adhoc / Control / Power save poll	40	7.2273
Adhoc / Control / RTS	45	7.2063
Adhoc / Control / CTS	41	7.3242
Adhoc / Data / Data	40	7.3333

3. Under management frame for the probe request (fig 2.a) with SNR 19 dB with respective EVM rms of 24.6124% the signal can be successfully transmitted while for probe response (fig 2.b) EVM rms i.e. 22.6645% followed with an increase in SNR that is 20dB. Comparatively for beacon (fig 1.c) a less SNR (15dB) can give successful reception with a very high value of EVM rms reached (52.6606%). i.e. a large error can be tolerated with maintenance of signal power in between 15 to 20 dB. Similarly comparing the Authentication (fig 2.d) and Deauthentication (fig 2.e) subtype. The deauthentication require large signal power for somewhat similar EVM rms fall. That is to connect a call with reference to fig 2.a, 2.b, 2.c a signal strength between 15 dB to 20 dB required that is feasible variation but EVMrms is getting to large due to the impairments.

4. As we gone through the management frame the max strength required noticed is for deauthentication frame (fig 2.e) in comparative study of 5 subtypes considered while it can be observed with the simulation result of control frame that the minimum SNR required is 40dB while there is great decrement in EVMrms in comparison of management frame. Here maximum it is only 7.3242 i.e for 'clear to send' (fig 2.c) subtype.

5. The DATA subtype used for the purpose of just transmitting the frame body or data from 802.11bg transmitter to 802.11bg receiver. As per simulation result analysis it require a high SNR of 40dB while the respective EVMrms recorded is 7.3333% i.e by maintaining a low EVMrms it requires a tough maintained SNR of 40dB (fig 4.a).

8.CONCLUSION

EVM should be minimized in order to enhance the performance of data networks. So it could be better parameter to analyse the desired QoS. In the work it is observed that different frame-type (Management, Control or data) responds differently to AWGN enabled channel impairment i.e. require different power level for maintaining QoS.

1. Management frames can be paged at a lower power level with the possibility of large EVMrms (24.612 %) Fig. 2.a

2. The control and data frames require large signal power to over come the hostile fading environment by Keeping Eb/No above 40 dB.

3. Control and data frames are very susceptible to fading environment and can not tolerate Modulation error beyond 7 %.

4. Adhoc network require a large signal strength for successful transmission of a signal frame and minimum possible EVMrms is 7.20633. So if we are capable of maintaining a power level between 15dB to 40 dB.

5. Measurement report suggest that EVM is a better candidate to measure Quality of Service of a given network. EVM obtained from different Networks in the proximity of Mobile node will help Mobile node to connect to a new target network bearing better QoS. This report is also helpful during a Network failure or poor QoS offered by existing Service provider or network.

REFERENCES

- [1] Rajender kumar, Brahmjit singh, "Performance characteristic measurements of Management frames" in IEEE 802.11x for VHO decision, WMIC, IEEE sponsored, PP. 210, January 2008.
- [2] E. Setton, T. Yoo, X. Zhu, A. Goldsmith and B. Girod, "Cross-layer Design of Ad Hoc Networks for Real-Time Video Streaming", IEEE Wireless Communications Magazine, vol. 12, no. 4, pp. 59-65, August 2005.
- [3] Q. Song and A. Jamalipour, "A Network Selection Mechanism for Next Generation Networks," in Proc. IEEE ICC'05, Seoul, Korea, May 2005.
- [4] J. Pinto, I. Darwazeh, "Simulation of a Base Station Receiver for the GSM Evolution -EDGE", Proceedings of International Conference on Telecommunications. ICT'2000, Vol. 1, Acapulco, Mexico, 22-25 May 2000, pp. 121-125.
- [5] R. Hassun, M. Flaherty, R. Matrieci, M. Taylor, "Effective Evaluation of Link Quality using Error Vector Magnitude Techniques", Proceedings of 1997 Wireless Communications Conference, Boulder, CO, USA 11-13 August 1997, pp. 89-94.
- [6] W. Chen, J. Liu, and H. Huang, "An Adaptive Scheme for Vertical Handoff in Wireless Overlay Networks," in Proc. of ICPADS'04, Newport Beach, CA, July 2004.
- [7] F. Zhu and J. MacNair, "Optimizations for Vertical Handoff Decision Algorithms," in Proc. IEEE WCNC'04, Atlanta, GA, March 2004.
- [8] 3GPP, "QoS Concepts and Architecture," TS 22.107 (v6.3.0), June 2005.
- [9] Hewlett-Packard Co., Notes on Error Vector Measurements with the HP894000 Vector Signal Analyzer. Jan 1996.
- [10]. www.seasolve.com, for WiLANTA software

Association Rule in Web Usage mining

Prof. Marathe Dagadu Mitharam

ABSTRACT:-

Association Rule is a fundamental of Data mining task. Its objective to find all co-occurrence relationship called, Association among data item. Let $I = \{i1, i2, \dots, im\}$ be a set of **items**. Let $T = (t1, t2, \dots, tn)$ be a set of **transactions**.

Goal: - Association and Correlation Analysis

In this paper describes

- 1) Discovery and analysis of web usage patterns using Association analysis.

* * * _ _ * * *

1) INTRODUCTION:-

Association rule discovery and statistical correlation analysis can find groups of items or pages that are commonly accessed or purchased together. Association based on Apriori algorithm. This algorithm finds groups of item using support and confidence. Satisfying a user specified minimum support threshold). Such groups of items are referred to as frequent itemsets & frequent itemsets graph.

Log files generated by web servers contain enormous amounts of web usage data that is potentially valuable for understanding the behavior of website visitors.

e.g.

$$\text{Support} = \frac{(X \cup Y).Count}{n}$$

$$\text{Confidence} = \frac{(X \cup Y).Count}{X.Count}$$

2) SUPPORT & CONFIDENCE:-

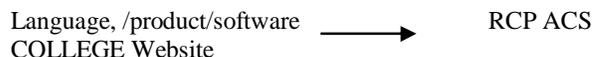
The Support of rule, $X \rightarrow Y$ the percentage of transaction in T that contains $X \cup Y$. n is the number of transaction in T. Support is useful measurement of itemset or items. If X is true then checks for Y, if X is false then nothing to be say Y. In the following example X union Y then count.

Using above examples we can accept the **minsub** and **minconf**. To calculate minsub and minconf as follows.

T1	C++, JAVA, RUBY
T2	C++, ASP
T3	ASP, VB
T4	C++, JAVA, ASP
T5	C++, JAVA, PHP, ASP, RUBY
T6	JAVA, PHP, RUBY
T7	JAVA, RUBY, PHP
JAVA, PHP	RUBY \longrightarrow [sup=3/7,conf=3/3]

In above 7 transactions JAVA, PHP & RUBY show 3/7 times. Every item checks itemset to every using **Joining** and **Pruning** steps. In web usage mining such rule can be use to optimize structure of website.

e.g.



3) EXPERIMENT- FINDING WEB USAGE ASSOCIATION RULES:-

Relation: weather					
No.	outlook Nominal	temperature Numeric	humidity Numeric	windy Nominal	play Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy	70.0	96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0	70.0	TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	no
9	sunny	69.0	70.0	FALSE	yes
10	rainy	75.0	80.0	FALSE	yes
11	sunny	75.0	70.0	TRUE	yes
12	overcast	72.0	90.0	TRUE	yes
13	overcast	81.0	75.0	FALSE	yes
14	rainy	71.0	91.0	TRUE	no

Instances: 14

Attributes: 5

- outlook
- temperature
- humidity
- windy
- play

The purpose of this experiment was to give some insight into the usefulness of association rules when they are applied to the web log data set of an education institution and others. We expected to find rules that correlate to web pages that contain information about sunny, rainy or temperature etc.

If check sunny, false → yes [sub 1/14, conf 1/1]

Suppose this is transaction table and find out Frequent Itemset then,

T1	C++, JAVA, RUBY	T2	C++, ASP
T3	ASP, VB	T4	C++, JAVA, ASP
T5	C++, JAVA, PHP, ASP, RUBY	T6	JAVA, PHP, RUBY
T7	JAVA, RUBY, PHP		

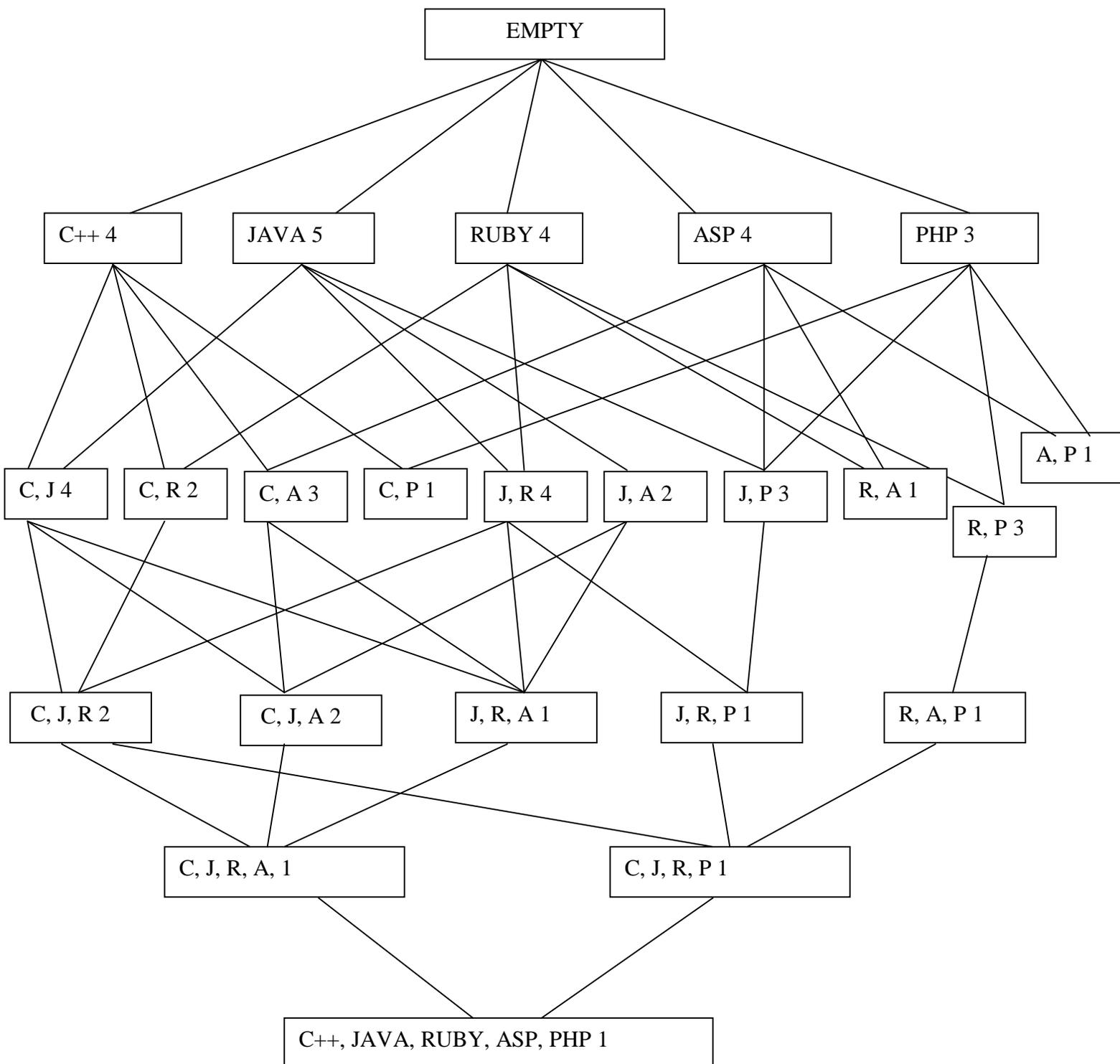
Size1		Size 2		Size 3		Size 4	
Item set	Supp.	Item set	Supp.	Item set	Supp.	Item set	Supp.
C++	4	C++, JAVA	3	C++, JAVA, RUBY	2	C++, JAVA, RUBY, ASP	1
JAVA	5	C++, RUBY	2	C++, JAVA, ASP	2	C++, JAVA, RUBY, PHP	1
RUBY	4	C++, ASP	3	JAVA, RUBY, ASP	1		
ASP	4	C++, PHP	1	JAVA, RUBY, PHP	3		
VB	1	JAVA, RUBY	4	RUBY, ASP, PHP	1		
PHP	3	JAVA, ASP	2				
		JAVA, PHP	3				
		RUBY, ASP	1				
		RUBY, PHP	3				
		ASP, PHP	1				

Fig. 1 Web transactions and resulting frequent itemsets (minsup = 1)
In the Fig 1 find out frequent itemset by using joining and pruning methods of association rule.

4) FREQUENT ITEMSET GRAPH:-

Finds items *C++* and *RUBY* as candidate recommendations. The recommendation scores of item *A* and *C* are 1, corresponding to the confidences of the rules, *JAVA, ASP ->C++* and *JAVA, ASP->RUBY*, respectively.

A problem with using a single global minimum support threshold in association rule mining is that the discovered patterns will not include “rare” but important items which may not occur frequently in the transaction data.



C=C++ **J=JAVA** **A=ASP**
R=RUBY **P=PHP**

Fig 2. Frequent itemset

5) CONCLUSION:-

This paper has attempted to for the purpose of web usage mining. The proposed methods were successfully tested on the log files using association rule.

We conducted a comprehensive analysis of web usage association rules found in a website of an educational institution. Our experiments confirmed that one of the major issues in association rule finding is the existence of too many rules, all of which satisfy defined constraints, but it is difficult to exploit and identify those that are truly interesting to the user.

Using this paper the user easily solved the support, confidence and frequent itemset in server log file.

6) REFERNECES:-

- 1) Web data mining – Bing Liu
- 2) PPT for Association Rule- Bing Liu

3) Srivastava, J., Cooley, R., Deshpande, M., Tan, P.N.(2000). Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. ACM SIGKDD, Jan 2000.

4) Jaideep Srivastava Paper

5) WCA. Web characterization terminology & definitions.

6) <http://www.w3.org/1999/05/WCA-terms/>. Vigente al 19/11/2005

Author Name:- Prof. Marathe Dagadu Mitharam
R.C.Patel A.C.S. College, Shipur
At/Post- Thalner Tal- Shirpur Dist- Dhule
(Maharashtra)

Slotted Plus Shape Microstrip Antenna with enhanced bandwidth

Jagadeesha.S, Vani R.M, P.V. Hunagund

Abstract: In this paper a concept of broadband slotted plus shape microstrip antenna with open end meandering slots in the ground plane is presented. Three identical narrow open end meandering slots were embedded in the antenna ground plane parallel to plus shaped radiating edge and study is made. This antenna is radiating at dual resonant frequencies This covers part of wibro(2300-2390), full band of ISM (2400-2484MHz), Full band of satellite DMB(2605-2655MHz) and overall bandwidth of 474 MHz(19.1%) are achieved. The slotted plus shaped microstrip antenna is simulated using the method of moments based commercial software (IE3D) and was found to perform well in terms of bandwidth and , radiation patterns.

Keywords: Microstrip antenna, multi band antenna. Wide band antenna, Wideband application, slot antenna, plus shaped antenna, Ground plane based antenna

◆

1. Introduction

Rapid progress in wireless communication promises to replace wired communication networks in the near future in which antennas plays a more important role. Microstrip antennas are used in a broad range of applications from communication systems to biomedical system, primarily due to several attractive properties such as light weight, low profile, low production cost, conformability, reproducibility, reliability, and ease in fabrication and integration with solid state devices. One of the most serious disadvantages of microstrip antenna is their limited band width. Researchers investigated method to increase single bandwidth [1-5]. Meandering technique when applied to ground plane of the microstrip antenna was proved to be one of effective method in reducing the size of the microstrip antenna and enhancement of bandwidth[6]. However, the obtained bandwidth in this case is comparatively very less when met with the requirements of the above applications. Further enhancement in the antenna bandwidth and size reduction was very much in need. Many combinations of radiating patch and ground plane slots were configured and analyzed to achieve extreme compact and broad band antennas. T.W Chiou et al [7] introducing slots in the ground plane that helped to enhance bandwidth of a terminal antenna at the upper end of the invested frequency band. Recently broadening of the lower operating frequency band was achieved using two open end slots in the ground plane under the radiating element(inverted F-Typed patch)[8]. It is observed that the use of two parallel open end slots in the ground plane

which results in enhanced band width there by generating dual frequencies. In this paper we present slotted plus shape microstrip antenna with open end meandering slots in the ground plane. Three identical narrow open end slots were embedded in the plus shaped antenna ground plane. These slots are aligned with an equal spacing of $L/4$ parallel to radiating edge.

2. Design Consideration

Plus shaped patch antenna which is considered as base antenna whose two strips of dimensions are considered 15.1mm *45.3mm & 11.8mm*35.4mm respectively is mounted on dielectric substrate of thickness 1.6mm and material used is Glass epoxy with relative dielectric of $\epsilon_r=4.4$ is designed for operating frequency of 2.2GHz. This rectangular patch in turn fed by center fed microstrip line feed of dimensions(L f50, W f50)=18.4mm,3.05mm through a quarter wave transformer having (Lt50 , Wt50)=18.55mm, 0.5mm). They are mounted on a ground plane of dimension 82mm*55mm fed by microstrip line using 50 ohm SMA connector for wideband applications is shown in figure 1a. Study has been made further by considering base shape of antenna by inserting horizontal slots on both side with respect to center of patch. Optimized distance between slots is being considered as 2mm which gives lowest frequency but reduced band which is represented in figure 1b. In order to enhance band width later we consider three identical narrow open end slots were embedded in the antenna's ground plane that are aligned with an equal spacing of $L/4$ parallel to radiating edge of the plus shaped patch.. The embedded open end slots are narrow (Ws=2mm) and have a slot length Ls of about 75% is considered as shown in figure1c. Above antenna gives maximum bandwidth.

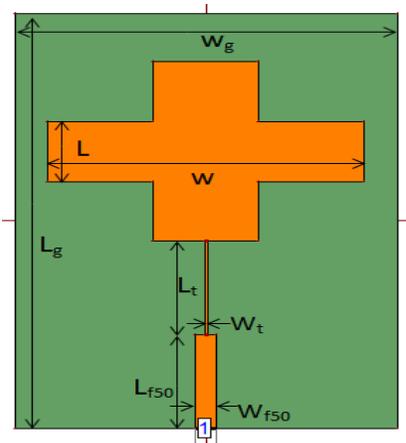


Figure1(a). Plus shaped base antenna

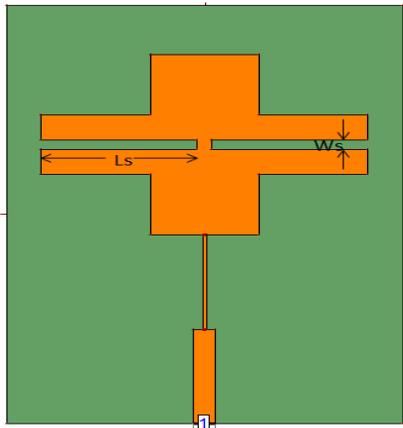


Figure1 (b). Geometry of base Antenna with slot in plus Shaped Patch.

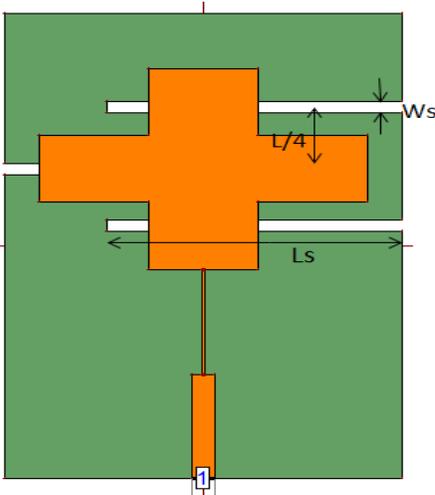


Figure1(c): Geometry of base Antenna with open end Meandering slots in the ground plane

3. Results and Discussions

The proposed Antenna design has been simulated using Zeland's simulation package i.e., IE3D. Comparative study has been made with respect to plus shaped base antenna without having slot in ground plane and by considering optimized slot on plus shape patch with distance between the slot as 2mm which is called Antenna1. It is also being compared with antenna 2 is having slot in ground plane. All the data is summarized in Table 1. Figure 2(a) shows the variation of return loss with frequency for base shape antenna. Variation of Return loss versus frequency for antenna 2 is shown in fig 2 (b). Similarly results for the effect of slot in ground plane is as shown in figure 2(c). From the results it is clear that the bandwidths of proposed antenna with slots in ground plane is more compared to other antennas. The antenna 2 also gives dual bands with overall bandwidth of 474MHz. Radiation patterns for plus shaped base and plus shaped patch antenna with slotted ground plane as shown in fig 3(a) & 3(b) respectively. From the figure3 (a) it is clear that low back lobe radiation is an added advantage for using this antenna in a cellular phone, since it reduces the amount of electromagnetic radiation which travel towards the user head.

Prototype Antenna	Resonant frequency	Return loss	Band width	Over all band width
Plus shaped base antenna	$f_1=2.2$ $f_2=2.49$	-11.63 -11.64	20 100	120
Base antenna with slot(Antenna 1)	$f_1=2.57$	-11	190	190
Base antenna with slot in Ground plane(Antenna2)	$f_1=2.25$ $f_2=2.61$	-21 -26.5	184 290	474

Table1: The Simulated results of Proposed Antennas.

Figure3b: Graph for return loss v/s frequency open end meandering slots in the ground plane of plus shape antenna.

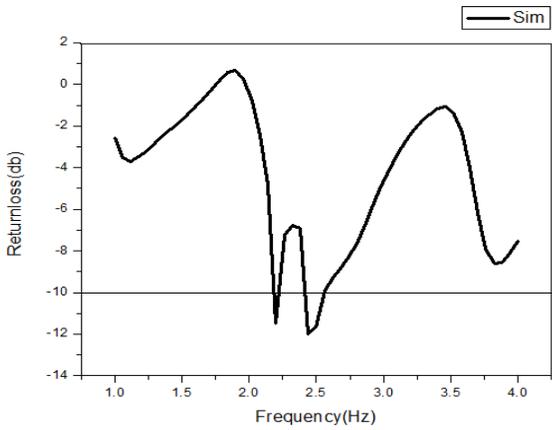


Fig 2a Graph for return loss v/s frequency for base shape

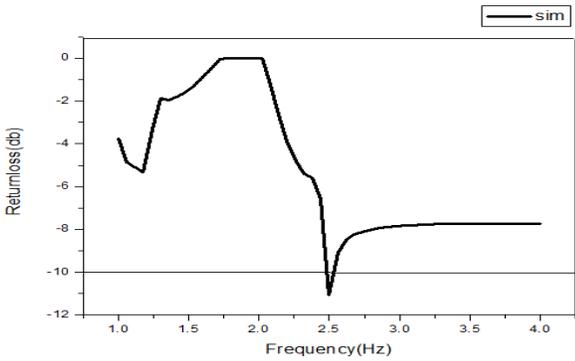


Fig 2b: Graph for return loss v/s frequency for slotted plus antenna.

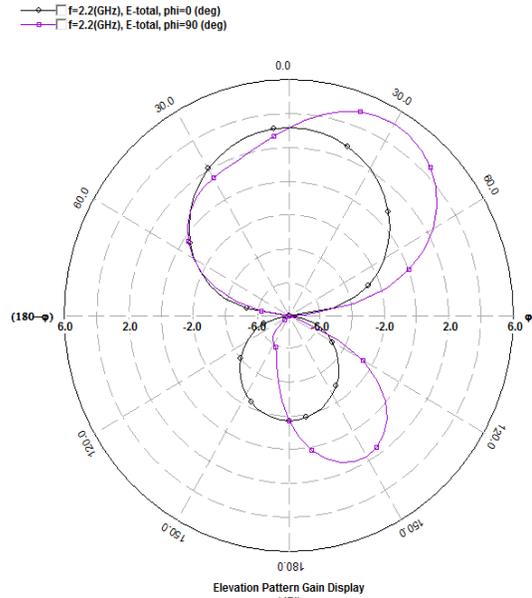
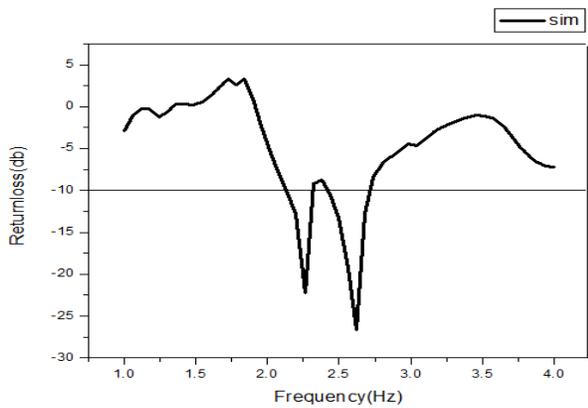


Figure3a: Radiation patterns of plus shaped base at 2.2 GHz.

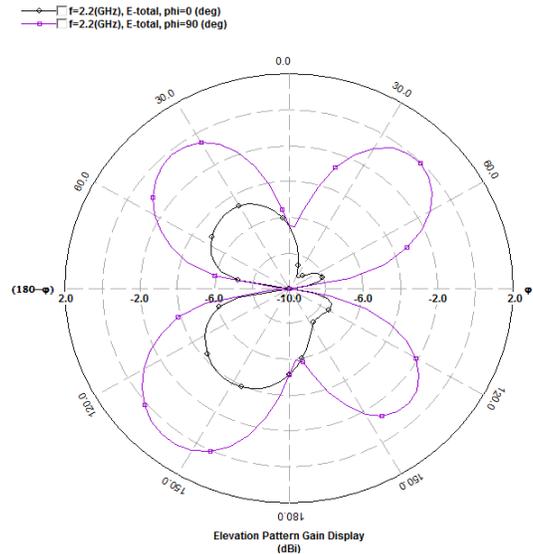


Figure 3b: Radiation pattern of slotted ground plane based plus shaped antenna at 2.25GHz.

4. Conclusion

This paper outlines slotted ground plane based plus shape antenna which gives total band width of 474MHz (19.1%). This bandwidth satisfy the requirements of the many services including part of wibro, full band of ISM and full band of satellite DMB applications. So from the results we conclude that the modified antenna with slots in ground plane gives enhanced bandwidth compared to other configurations.

Acknowledgement

The authors acknowledge their thanks to UGC, New Delhi for sanctioning the IE3D software under major research project which is most useful and reliable for designing microstrip antennas.

References:

[1] Garg.Ramesh, et al., Microstrip Antenna design Hand book, 2001 Artech house, Inc.

[2] A.K. Skrivervik, J.F.Aurcher, O staub and J.R. Mosig., PCS Antenna Design: The challenge of miniaturization, IEEE Antennas and propagation Magazine, Vol.43, Issue 4, PP12-27, Aug-2001.

[3] D.Sing, P, Gardner and P.S. Hall., Miniaturized microstrip antenna for MMIC Applications, Electronic Letters. Vol.33, No22, PP-1930-1, October1997.

[4] Bendad, N and K sarbndi, "A Multiresonant single element wideband slot antenna" IEEE Trans. Antennas & wireless propagation let., Vol 3, 5-8, 2004.

[5] Size, J-Y and K-L.Wing, "Bandwidth enhancement of a microstrip-line fed printed wide slot antenna" IEEE Trans. Antennas and propagate, Vol49, No7, 1020-1024, July 2001

[6] J.S.Kuo and K.L. Wong, "A compact microstrip antenna with meandering slots in the ground plane, "Microwave opt.Technol. Vol29, PP, 95-97, 2001.

[7].T.W. Chiou and K.L. Wong, "Design of compact microstrip antenna with a slotted ground plane", IEEE APS. Int. Symp. Dig. , PP 732-735, 2001.

[8] R.Hossa, A.Byndas and M.E. Bialkowski, "Improvement of compact terminal antenna performance by incorporating open-end slots in the ground plane", IEEE Microwave and wireless commn letts, vol. 14, June 2004.

- Jagadeesha.S, M.E., S.D.M Institute of Technology, Ujire-574 240, Mangalore (D.K), Karnataka, India, Mobile: +91 9844294915 E-mail: Jagadeesh.sd69@gmail.com.
- Vani.R.M., M.Tech., Ph.D., University Science Instrumentation Center, Gulbarga University, India. Phone: +91 08472263259, E-mail: vanirm12@rediffmail.com.
- P.V. Hunagund, M.Sc., Ph.D., Dept. of PG Studies and Research in Applied electronics, Gulbarga University, India, Mobile: +91 8472246029 E-mail: prabhakar_hunagund@yahoo.com.

Analysis and Behavioral Study of Substrate Noise Coupling in Silicon Integrated Circuits

Mandeep Kaur, Rajesh Mehra

Abstract- In this paper the behavioral study is presented on uniformly doped silicon substrate. Resistivity and thickness of substrate is varied and analyzed. It is observed that with the increase in resistivity and decrease in thickness of substrate, crosstalk problem existing in mixed circuits is reduced. FEM technology is used and HFSS simulation is done for S parameter extraction for $1000 \times 1000 \mu\text{m}^2$ sized substrate with two square contacts of size $50 \times 50 \mu\text{m}^2$, where 't' refers to thickness of substrate.

Key Words: Substrate, substrate noise, coupling, finite element method, doping, resistivity, parasitic elements, isolation, HFSS.

◆

1 INTRODUCTION

Single chip mixed signal designs combining digital and RF/analog blocks built over a common substrate provide reduced levels of cost, power dissipation and smaller package count. However its disadvantage is that the important fault introduced is substrate noise. Substrate noise isolation is getting more and more difficult to handle with new and upcoming technologies. With the increase level of integration in IC's today and increase in digital circuits speed, the problem of substrate is becoming more and more prominent. The performance of overall chip degrades. This problem is observed during testing phase only after chip is fabricated. Substrate noise determination can reduce the turnaround time for circuits and prevent costly design. In a lightly doped substrate, this analysis of the substrate noise problem is divided into three aspects [1], i.e., generation, propagation, and impact of substrate noise. In this paper a behavioral study of coupling in a common silicon substrate is presented. This paper also shows dependence of cross talk on substrate contact separation, substrate thickness. Substrate resistivity and signal frequency.

2 FINITE ELEMENT METHOD

The finite element method subdivides the entire domain into triangular (2D) or tetrahedral (3D) elements and

treats the elements as linear. The resulting linear relations between the potentials at the corners of the elements (the FEM nodes) can then be expressed by a piecewise linear (or piecewise planar for 3D) function. This linear function implies that every edge in the FEM discretization represents a resistance. In other words, the FEM discretization is equivalent to a resistance network. The resulting network is large and sparse. However, only a small number of the FEM nodes are actually terminal nodes, while the rest of the FEM nodes are internal nodes. Only the terminal nodes are connected to other physical structures (e.g. transistors, vias or possibly a parasitic interconnect capacitance). Therefore, the FEM network will have to be solved such that the internal nodes are eliminated and only the resistances between the terminal nodes remain. As such, the FEM can be very accurate and flexible. Even though it is possible to use the FEM in such cases, its limited speed might be prohibitive, especially for larger-scale modeling problems. A substrate is a medium composed of several conductive layers. A deposition process is done using appropriate materials on the top of these layers. The ports or contacts correspond to lightly doped but strongly conductive areas [8].

3 HFSS

"HFSS is the industry-standard software for S-parameter, full-wave SPICE extraction and

electromagnetic simulation of high-frequency and high-speed components. HFSS is widely used for the design of on-chip embedded passives, PCB interconnects, antennas, RF/microwave components, and high-frequency IC packages. HFSS improves engineering productivity, reduces development time, and better assures first-pass design success.

The latest release of HFSS delivers significant productivity gains to Microwave/RF engineers and expands electromagnetic co-design to a new segment of engineers working in the areas of RF/analog IC and multi-gigabit designs as well as EMI/EMC. HFSS uses the Finite Element Method (FEM) to solve Maxwell's equations.

The primary advantage of the FEM for solving partial differential equations lies in the ability of the basic building blocks used to discretize the model to conform to arbitrary geometry. The arbitrary shape of the basic building block (tetrahedron) as shown in Fig.1 also allows HFSS to generate a coarse mesh where fewer cells are needed to yield an accurate solution, while creating a finely discretized mesh where the field is rapidly varying or higher accuracy is needed to obtain an accurate global solution.

The FEM has been a standard for solving electromagnetic problems since the inception of HFSS in 1990. The FEM has been a standard for solving problems in structure mechanics since the mid 1950's.

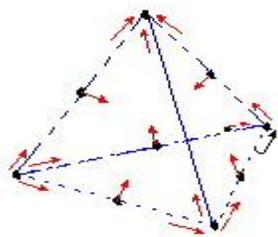


Fig.1 Tetrahedron shape of mesh

Substrate coupling occurs in integrated circuits due to the conductive and capacitive path which exists in the silicon substrate. Conductive path of silicon substrate is predominant at low frequencies, while the capacitive

path becomes significant at high frequencies. Noise is injected into the substrate through resistive injection (e.g. impact ionization current and ohmic guarding) and capacitive injection (e.g. junction capacitance). Substrate noise reception mainly takes place through capacitive sensing (i.e. junction capacitances). Substrate noise, usually generated by noisy circuit block (i.e. digital block), received by sensitive analog/RF blocks can severely degrade its performance, thus degrades the performance of the overall system. There are quite a few substrate crosstalk suppression techniques that have been widely implemented on industrial level. Guard ring is one of the most common techniques used, where p^+ silicon substrate ring provides low impedance return path for the substrate noise towards the ground. Similarly, buried layer and triple-well isolation are used in BiCMOS and CMOS processes, respectively. Instead of using low resistivity substrate, a piece of silicon substrate can be modeled by an RC network, where the R and C depend on the substrate resistivity and dielectric constant, respectively. While the capacitance takes effect at relatively high frequencies, the resistance is dominant at low frequencies. The high-resistivity substrate can improve the isolation below the cut-off frequency but starts to lose its isolation effectiveness at earlier stage which makes it less suitable for high-frequency application with respect to isolation, not to mention the higher cost of high-resistivity silicon substrates.

4 MODELING OF SUBSTRATE COUPLING

The numerical modeling using the device solver requires considerable computational time. Therefore, it is desirable to build an equivalent circuit model comprised of ideal lumped elements. The equivalent circuit can be solved using common circuit simulators (e.g. SPICE, ADS). Modeling substrate noise coupling is very complex since it includes all noise injectors and receptors built on a particular substrate. Substrate doping can vary from 1S/m to 10000S/m. 10S/m is for lightly doped substrate and 10000S/m for heavily doped

substrate. The cut off frequency $f_c = \frac{1}{2\pi\rho_{sub}\epsilon_{Si}}$ It is 15Ghz for lightly doped substrates .So frequency<15Ghz substrate behave resistively and capacitance is prominent at frequencies >15Ghz [4]. Frequency vs Resistivity curve is shown in Fig.2, which shows that frequency is inversely proportional to resistivity of substrate

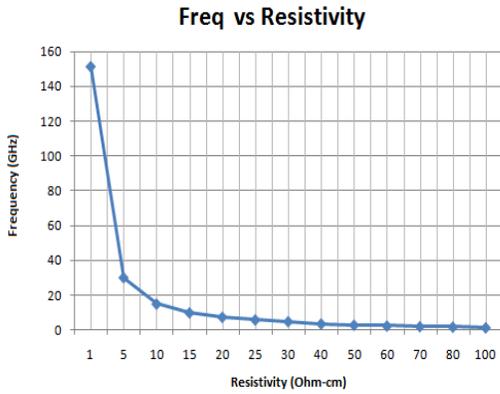


Fig.2 Frequency vs resistivity graph

But it is clear that on such substrates reactance cannot be ignored with RF frequencies. For higher frequencies inductive skin effect and slow wave effect can rise. Here we have considered uniformly distributed substrat.This paper is limited to behavioral analysis of a substrate with small number of contacts i.e. two. Contact areas are where circuit interacts with substrate. Equipotential assumption is valid until contact is not too large. Contacts interact without direct coupling but with resistance they are still indirectly coupled. The device to substrate junctions are considered as equipotential contacts with surface. Substrate has distributed impedance. Substrate provides a parasitic coupling path for signals within same circuit. Highly mixed circuits provide crosstalk.

In this paper, model consist of two metal substrate contacts of size $50 \times 50 \mu\text{m}^2$ and silicon substrate of size $1000 \times 1000 \mu\text{m}^2$ with variable thickness of 50, 100, 500, 1000 μm as shown in Fig.3, where d is the distance between two contacts, s is the size of contact and S_{sub} is the size of substrate.

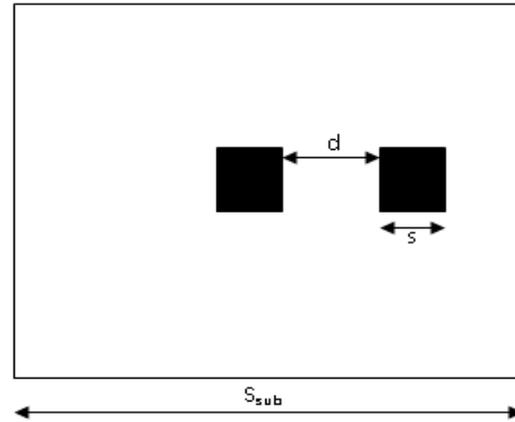


Fig.3 Substrate model

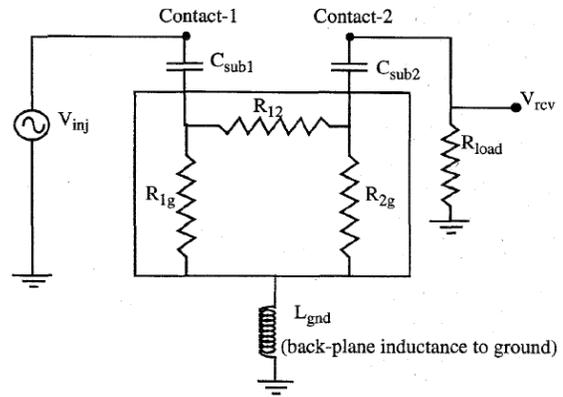


Fig.4 Equivalent circuit model

Fig.4 represent equivalent circuit model where R_{12} is resistance between two contacts, R_{1g} and R_{2g} are resistance between ground and contact [2].

HFSS use FEM to solve the field equations in a volume. It does not solve carrier continuity equations. The substrate coupling is represented in terms of isolation between contacts [5]. This isolation depends on several factors besides substrate model elements. These include contact to substrate capacitance, frequency of operation, load impedance connected to receiver contact and inductance of bond wire connected to gnd. the complete model shown in fig4can be simulated using ADS. Contact pads are sealed as 2D objects. They are defined with finite conductivity boundaries. There is air box which enclose the substrate has radiation boundary condition which means that all waves will be absorbed. Here C_{sub1} is equal to C_{sub2} due to symmetry and represents the contact to substrate capacitance of contact

pads. R_{load} is the load resistance; L_{gnd} is the bond wire inductance is needed in order to provide an accurate model of real structure. R_{1g} and R_{2g} have same values as structure is symmetrical. Isolation of test structure is defined as:

$$I = 20 \log (V_{rev}/V_{inj}) \text{ dB} [2] \quad (1)$$

5 RESULTS AND DISCUSSIONS

The HFSS simulation has been done using FEM technology, which is faster and accurate than Substrum and Spice. The isolation parameter S21 has been simulated for different substrate thicknesses by varying distance between two contacts. The resultant graph is deduced. The thickness is varied from 50 to 1000 μm .

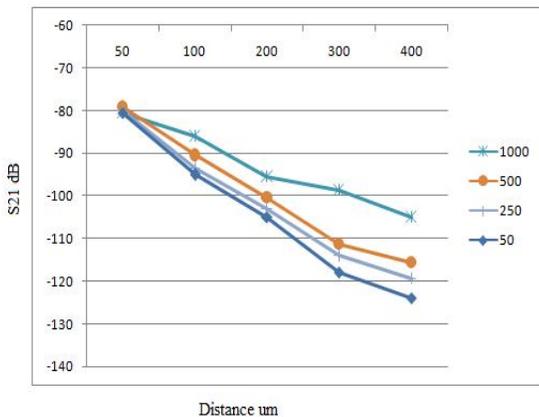


Fig.5 Isolation vs distance graph for different thickness

Fig.5 shows that with decrease in thickness there is more reduction in crosstalk.

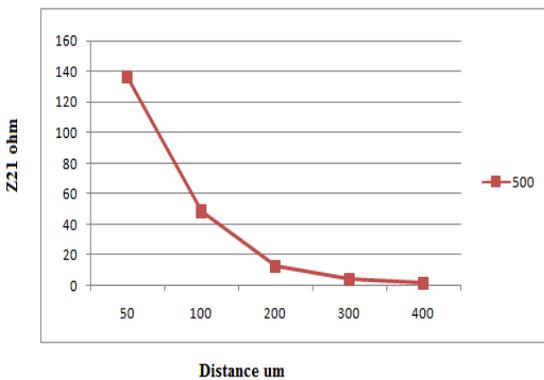


Fig.6 Impedance vs distance graph for thickness of 500 μm

The Z_{21} values are obtained from the calibrated layer thicknesses and resistivities. For noise estimation purpose the values of R_{1g} and R_{2g} are important, so these are calculated using the values of Z_{11} and Z_{21} [7], where

$$Z_{11} = R_{1g} \parallel (R_{12} + R_{2g}) \quad (2)$$

$$Z_{21} = (R_{1g} \times R_{2g}) / (R_{1g} + R_{12} + R_{2g}). \quad (3)$$

The Fig.6 shows that overall impedance decreases linearly at high frequencies with increase in distance between contacts. The impedance graph of substrate is evaluated for thickness 500 μm .

The resistance vs distance graph at different resistivities is shown in Fig.7 at 1 GHz frequency, which shows that for higher resistivity the resistance increases, with increase in distance.

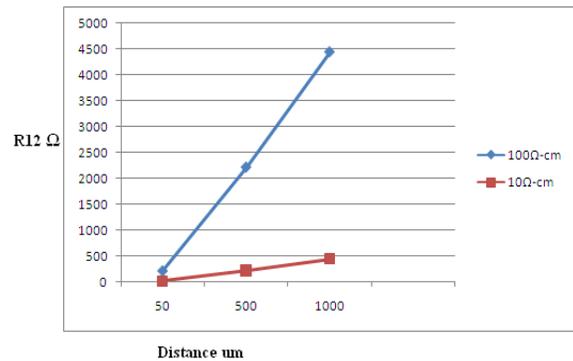


Fig.7 Resistance vs distance for different resistivities

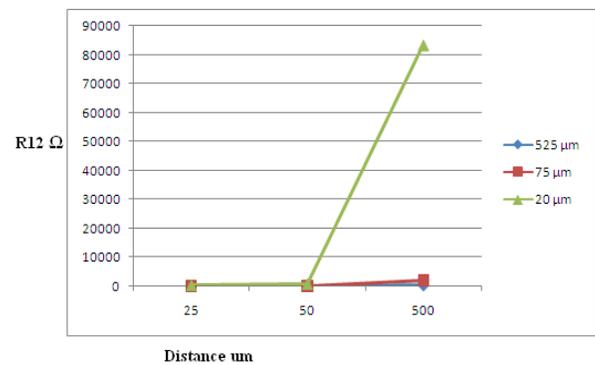


Fig.8 Resistance vs distance graph for different thicknesses

With more increase in distance, isolation starts becoming independent of distance. R_{12} increases more rapidly with thinner substrates as shown in Fig.8 for Si substrate with 100ohm-cm resistivity and thickness is varied from 20 μ m to 525 μ m. Isolation increases with thinner substrate and it also improves with increasing the substrate resistivity. Isolation is related to distance between contacts to certain extent Distance driven isolation saturates as distance gets larger than the areas of transmitter and receiver.

Nishath K. Verghese, Terri S. Fiez, Kartikeya Mayaram, "A Predictive Methodology for Accurate Substrate Parasitic Extraction", ICAS, Vol.5, pp.149-152, IEEE 2004
[8] N.Masoumi, M. I. Elmadry, S. Safavi-Naeini, H. Hadi, "A Novel Analytical for evaluation of Substrate Crosstalk in VLSI circuits", The First IEEE workshop on Electronic Design, Test and Application, pp.355-359, 2002

REFERENCES

- [1] S. Donnay and G. Gielen, "Substrate Noise Coupling in Mixed-Signal IC's", Norwell, MA: Kluwer, 2003.
- [2] Ranjit Gharpurey and Robert G. Meyer, "Modeling and Analysis of Substrate coupling in Integrated Circuits", IEEE JOURNAL OF SOLID-STATE CIRCUITS, vol.31, no3, pp.344-353 March 1996.
- [3] K.rajendran, J.B Johnson, S.Furkey, M.Kumar and S. Fischer, "Measurement and 3D Simulation of Substrate Noise Isolation and Resistance for Mixed Signal Applications", IEEE topical meeting on silicon monolithic integrated circuits in RF systems. 2004
- [4] R. Gharpurey and R.G. Meyer, "Modeling and analysis of substrate coupling in IC's", in Proc. IEEE Custom-Integrated Circuit Conf., pp. 125-128, May 1995.
- [5] Salman, E.; Jakushokas, R.; Friedman, E.G.; Secareanu, R.M.; Hartin, O.L, "Methodology for efficient substrate noise analysis in Large- Scale mixed signal circuit", IEEE Transactions on Very Large Scale Integration (VLSI) systems, Vol.17, pp.1405-1418, October 2009.
- [6] Emre Salman, "Switching Noise and Timing Characteristics in Nano-scale Integrated Circuits" Doctorate thesis, 2009
- [7] Ajit Sharma, Chenggang Xu, Wen Kung Chu,

-
- Mandeep Kaur is currently pursuing masters degree program in Electronics and Communication Engineering, E-mail: sandhu.mandeekaur@gmail.com
 - Rajesh Mehra is currently pursuing P.Hd program in Electronics and Communication Engineering

Pseudo Elastic Behaviour of Liquid Binary Alloys

Dr. A. K. Pandey, Dr. N. Alam, Dr. Ashok Kumar

Abstract- The trend in the elastic constants of simple and transition elemental cube metals are explained in terms of a uniform electron gas theory. Previously this new model was used to show that the cohesive energies of the elemental transition metals depend primarily on the bonding valence and the average electron density at the boundary of the unit cell. In the present work we shall study the pseudo elastic behavior of liquid binary alloys using pseudo potential model based on the density functional theory with both the local density approximation and the generalized gradient approximation for the exchange correlation function. Here we use a slight modification of the same model to show that the elastic constants of the elemental cubic metal depend primarily on the bonding valence, the density at the cell boundary and the symmetry of the lattice.

Index Terms- Elastic Constants, Cohesive Energy, Pseudopotential Model, Density Functional Theory, Generalized Gradient Approximation, Exchange Correlation Function, Bonding Valence

1 INTRODUCTION

THE effect of strain on electron properties require knowledge of the materials mechanical properties and specifically the elastic constants which describe the response to an applied macroscopic stress. To this purpose the elastic constants of the material under study are of particular interest and has been calculated at different composition by computing the compounds of the stress tensor δ for small strain. It is well known that a cubic crystal has only three independent elastic constants. When a vacancy is thermally created in a solid, the latter's volume changes thus defining a vacancy formation volume. In other cases a defect is produced by exchange one of the foreign atom or molecule. A defect volume V^d can be define as the difference between the volume V of the host material containing defect (d) and the volume V^o of the perfect crystal, provided that the volume V and V^o contain the same number of particles. The theoretical pressure volume relations for metals were studied by Soma et al. By using a model potential due to Kulshrestha et al. Pandya et al. have also studied the pressure volume relation for fcc transition metals using the model proposed by Antanov et al. Some rigorous methods to calculate the cubic and hexagonal elastic constants have been well described in previous studies.

The main purpose of the present work is the determination of the elastic constants of liquid binary alloys at different concentration using pseudopotential model by means of the Cambridge Serial total Energy Package Soft Ware(CASTEP) Employing the generalized gradient approximation within density functional theory (DFT).

2 Theoretical formalism

In the present work Varotsos and Alexopoulos model has been applied using slight modification in volume due to concentration.

Varotsos had derived an expression for the bulk modulus of mixed system in terms of constituent ionic compounds, their molar volumes and their molecular fractions. According to him the bulk modulus of the mixed system is given by

$$B = [1 + x (V_2/V_1 - 1)/1 + (B_1V_2/B_2V_1 - 1)] B_1$$

In this expression x is the molecular fraction of compound 1 having Bulk modulus and molar volume B_1 and V_1 respectively. B_2 and V_2 are the Bulk modulus and the molar volumes of compound 2 respectively.

The molecular volume of the alloy is determine in ideal sense using the relation

$$\Omega_{\text{alloy}}^{\text{ideal}} = (1 - C)\Omega_1 + C\Omega_2$$

Here Ω_1 and Ω_2 are the atomic volume of the compounds species of the alloy and C is concentration of the second compound.

With the increasing power of modern work stations pseudopotential calculation of elastic constants have become possible. These investigations are all based on density functional theory; differ however, in the treatment if tightly bound core electrons.

Our calculations were made using the plane wave pseudopotential method as invoked by the CASTEP. Castep is the total energy code that allows one to easily compute atomic forces and stress as well as the energy and electronic properties and it is thus most suitable for structural studies.

The Jellium model of the electron gas has long been used to understand the qualitative trends in the plasma ion and electron hole excitation spectra of elemental metals. The calculated spectra agree semiquantitatively with experiment for simple metals, and are qualitatively useful for the transition metals. One is used to thinking that the excitations of quantum systems are harder to predict than ground-state properties. From this point of view, it is surprising that the Jellium model has not provided a similarly simple explanation for the trends in the ground state energetic of the elemental metals. For non-uniform systems the

- P. G. Department of physics, Purnea College, Purnea Bihar (India) e-mail – akphysics01@gmail.com
- P. G. Department of physics, G. D. College, Begusarai Bihar (India) mob. No.- 9835273652
- P. G. Department of physics, B. N. Mandal Univ., Madhepura, Bihar (India)

background density varies with position and the new potential gives rise to force. The total external electron potential $V_{ext}(r)$ is given explicitly by

$$V_{ext}(r) = \phi(r) + v_0 n_b(r)/n, \quad (1)$$

Where $\phi(r)$ is the usual electrostatic potential. The strength of the ad hoc interaction v_0 is uniquely fixed by the condition of zero forces for the uniform state and is given in terms of the energy of bulk Jellium and n is the uniform equilibrium electron density.

The theory of ideal metals can be understood by starting with an electron gas that has a uniform electron and background density n . The zero-force condition is achieved by introducing an additional ad hoc electron-ion potential, which at a point r is proportional to the background density at that point. The strength of the additional potential is determined by the zero-force condition, and is given by

$$v_0 = -n(\partial e_{jell}/\partial n)_n \quad (2)$$

Here e_{jell} is the energy per electron in uniform Jellium of density n .

The Born-Oppenheimer approximation yields the following many-body Hamiltonian for the unmodified theory of ideal metals:

$$H = -\hbar^2/2m \sum_i \nabla_i^2 + e^2/2 \sum_{i \neq j} (1/r_{ij}) + e^2/2 \iint d^3r d^3r' [n_b(r) n_b(r') / r - r'] - e^2 \sum_i d^3r [n_b(r) / r_i - r_j] + v_0 \sum_i [n_b(r_i)/n] \quad (3)$$

A density function for the energy can also be used to describe the theory of ideal metals in the Born-Oppenheimer approximation. Formally,

$$E[n_e, n_b] = T_s[n_e] + \int d^3r \int d^3r' [\rho(r) \rho(r') / r - r'] + E_{xc}[n_e] + E_{ei}[n_e, n_b] \quad (4)$$

Here, $\rho = n_e - n_b$ denotes the net charge. E_{xc} denotes the exchange-correlation energy, while T_s denotes the kinetic energy of non-interacting electrons. The second term on right hand side denotes the classical electrostatics energy. Finally, the last term on the right hand side models electron-ion interactions beyond the classical electrostatic interaction. It is expressed, for the unmodified ideal metal as

$$E_{ei}[n_e, n_b] = v_0/n \int d^3r n_b(r) n_e(r) \quad (5)$$

We modify the theory of ideal metals by introducing the following, more general form for the electron ion interaction.

$$E_{ei}[n_e, n_b] \approx \int d^3r n_e(r) [v_0 \{n_b(r)/n\} \{1 - 1/2 \beta^2 (n_e(r) - n)^2\}] \quad (6)$$

Here β is an ad hoc parameter that is chosen once to give agreement with experiment for all metals.

Solution for the energy can be obtained from the energy-density function by following the procedure of Kohn and Sham. The self-consistent equations with an effective potential is therefore given as

$$v_{eff}(r) = \delta E_{xc} n_e / \delta n_e + \delta E_{ei} [n_e, n_b] / \delta n_e \quad (7)$$

Density-functional theory can also be used to compute the response of the electron gas to a perturbation. The screened potential-density linear response function can be obtained in the random-phase approximation Shore et al, χ

$$\chi^{sc}(q) \approx \chi^0(q) / [1 - (8\pi/q^2 + \delta^2 E_{xc} / \delta n_e^2) \chi^0(q) + \delta^2 E_{ei} / \delta n_e^2] \chi^0(q) \quad (8)$$

Here χ^0 is the Lindhard function. The functional derivative of the exchange-correlation energy can be evaluated in the local-density approximation, as

$$[\delta^2 E_{xc} / \delta n_e] \Big|_n = d^2/d n_e^2 \{n_e e_{xc}(n_e)\} \Big|_n \quad (9)$$

Here, e_{xc} denotes the exchange-correlation energy of Jellium per unit volume. The functional derivative for the electron-ion interaction can be evaluated from (6)

$$\delta^2 E_{ei} / \delta n_e \Big|_n = v_0 n_b/n d^2/d n_e^2 \{n_e [1 - 1/2 \beta^2 (n_e - n)^2]\} \quad (10)$$

If the transition metal is compressed, the rigid positive backgrounds of the various Wigner-Seitz cells overlap and the background becomes inhomogeneous, the Wigner-Seitz cell centered about the origin of coordinates and define the characteristic function as,

$$Y(r) = 1 \quad r \in \text{Wigner-Seitz cell}$$

and,

$$Y(r) = 0 \quad r \notin \text{Wigner-Seitz cell} \quad (11)$$

In the following, we will consider background densities that can be formed as

$$n_b(r) = n \sum_{R_i} Y(r - R_i) \quad (12)$$

Where R_i denotes the set of vectors.

Now the change in energy to second order $E^{(2)}$ is obtained by comparing three states. Which is

$$E_{eq}(n) = N [e_{jell}(n) + v_0] \quad (13)$$

Where N is the total number of electrons.

The extraction of elastic constants is straightforward once the energy changes have been calculated. The change in energy is related to the strain ϵ_{ij} and the elastic constants C_{ijkl} , via

$$\delta E = 1/2 C_{ijkl} \epsilon_{ij} \epsilon_{kl} \quad (14)$$

We computed the energy changes for the fcc and bcc cubic metals using following strains:

$$\begin{pmatrix} \delta & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} \delta & 0 & 0 \\ 0 & \delta & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} \delta & 0 & 0 \\ 0 & \delta & 0 \\ 0 & 0 & \delta \end{pmatrix} \& \begin{pmatrix} 0 & \delta & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (15)$$

The axes of strain tensor are aligned with the set of [100] direction of the cubic crystal. These strains over determine C_{11} , C_{12} , and C_{44} .

Vartsos, 1980; had derived an expression for bulk modulus of mixed systems in terms of those of constituent ionic compounds, their molar volumes, and their molecular fractions. This is given by

$$B = \left\{ \frac{1 + \eta (V_2/V_1 - 1)}{1 + \eta (B_1 V_2/B_2 V_1 - 1)} \right\} B_1 \quad (16)$$

In this expression η is the molecular fraction of compound 1 having bulk modulus and molar volume B_1 and V_1 , respectively. B_2 and V_2 are the bulk modulus and the molar volume of compound 2, respectively. B is the bulk modulus of the mixed system. This expression yields very satisfactory results. Moreover from the above expression we get the values for bulk moduli such, not the three independent elastic constants C_{11} , C_{12} , C_{44} . Which are sometimes necessary to enable one to predict the values of Young's modulus, Poisson's ratio, etc. in different directions.

The aim of this work is to derive expression to predict the values of C_{11} , C_{12} , C_{44} for mixed systems when those for constituent ionic compounds and their molecular fractions are known.

Giri and Mitra, 1985; have derived an expression for Debye temperatures $\Theta_{rs(pq)}$ of mixed system $rs(pq)$ in terms of those Θ_r and Θ_s of ionic compounds r and s having molecular fractions p and q respectively, which is

$$M_{rs(pq)} \Theta_{rs(pq)}^2 = p M_r \Theta_r^2 + q M_s \Theta_s^2 \quad (17)$$

Here $M_{rs(pq)} = p M_r + q M_s$, M_r and M_s are the molecular weights of the mixed system, compound r , and compound s , respectively. They have shown that the above equation is valid for number of mixed system at all proportions. For the cubic lattices the nearest neighbour central constant f_1 is related to Debye temperature Θ through the relation, Shirly, 1975;

$$f_1 = Y M \Theta^2 \quad (18)$$

Where,

$$Y = C k^2 A / 9 h^2 ;$$

A is the atomic mass unit, M the mass, C is a constant depending on the lattice type, having value 2.515 for f. c. c lattices.

If $f_1^{rs(pq)}$, f_1^r , f_1^s are the nearest neighbor central force constants for the mixed system, for compound r , and compound s , We may, therefore write

$$f_1^{rs(pq)} = Y M_{rs(pq)} \Theta_{rs(pq)}^2 \quad (19)$$

$$f_1^r = Y M_r \Theta_r^2 \quad (20)$$

$$f_1^s = Y M_s \Theta_s^2 \quad (21)$$

$$f_1^{rs(pq)} = p f_1^r + q f_1^s \quad (22)$$

Similarly,

$$f_2^{rs(pq)} = p f_2^r + q f_2^s \quad (23)$$

$$f_3^{rs(pq)} = p f_3^r + q f_3^s \quad (24)$$

So we get three equations for three types of force constants for mixed systems in terms for these constituent ionic compounds.

According to Niu and Shimizu, force constants f_1 , f_2 and f_3 are related to elastic constants C_{11} , C_{12} and C_{44} for f.c.c lattice through these three equations. They are

$$f_1 + 3 f_3 = a C_{44} \quad (25)$$

$$f_1 + 4 f_2 - f_3 = a (C_{11} - C_{44}) \quad (26)$$

$$2 (f_1 - f_3) = a (C_{12} + C_{44}) \quad (27)$$

Here a is the lattice constant. For mixed system it can be expressed as

$$a_{rs(pq)} C_{44}^{rs(pq)} = f_1^{rs(pq)} + 3 f_3^{rs(pq)} \quad (28)$$

$$a_{rs(pq)} C_{44}^{rs(pq)} = p f_1^r + q f_1^s + 3 p f_3^r + 3 q f_3^s = p (f_1^r + 3 f_3^r) + q (f_1^s + 3 f_3^s). \quad (29)$$

Finally,

$$a_{rs(pq)} C_{44}^{rs(pq)} = p a_r C_{44}^r - q a_s C_{44}^s \quad (30)$$

C_{44}^r and C_{44}^s are the values of C_{44} for ionic compounds r and s having lattice constants a_r and a_s respectively. $C_{44}^{rs(pq)}$ is that of the mixed system. $a_{rs(pq)}$ is the lattice constant of the mixed system with similar calculation for C_{11} and C_{12} we get

$$a_{rs(pq)} C_{11}^{rs(pq)} = p a_r C_{11}^r - q a_s C_{11}^s \quad (31)$$

$$a_{rs(pq)} C_{12}^{rs(pq)} = p a_r C_{12}^r - q a_s C_{12}^s \quad (32)$$

Thus we can derive C_{11} , C_{12} and C_{44} for mixed systems, if the constituent ionic compounds along with their lattice constants are known. These parameters of the reference binary alloys are not easily available, so we have used Vegard's law. Once by knowing the value of C_{11} , C_{12} and C_{44} , the value of B can be calculated.

3 Results and discussion

In the present work we have calculated the values of C_{11} , C_{12} and C_{44} for the reference alloy system at different concentration, which are tabulated in the table 1 – 3.

TABLE – 1

ELASTIC CONSTANTS OF Na–K AT EQUI-CONCENTRATION

q / 2k _r	Values of elastic constants (10 ¹⁰ N/m ²)			Values of bulk Modulii B (10 ¹⁰ N/m ²)	
	C ₁₁	C ₁₂	C ₄₄	Calc. (Present.)	Calc.(1.)
0.00	3.962	0.687	0.607	1.782	1.766
0.50	3.822	0.657	0.581	1.716	1.692
1.00	3.702	0.632	0.554	1.654	1.631
1.50	3.578	0.612	0.531	1.596	1.593
2.00	3.432	0.601	0.522	1.573	1.542

TABLE – 2

ELASTIC CONSTANTS OF K–Cl AT
EQUI–CONCENTRATION

q / 2k _f	Values of elastic constants (10 ¹⁰ N/m ²)			Values of bulk Modulii B (10 ¹⁰ N/m ²)	
	C ₁₁	C ₁₂	C ₄₄	Calc. (Present.)	Calc.(1.)
0.00	14.201	6.354	7.234	8.890	
0.50	13.365	6.001	6.935	8.732	
1.00	12.234	5.932	6.003	8.695	
1.50	11.759	5.634	5.443	8.235	
2.00	11.663	5.320	5.213	8.125	

TABLE – 3

ELASTIC CONSTANTS OF K–Br AT
EQUI–CONCENTRATION

q / 2k _f	Values of elastic constants (10 ¹⁰ N/m ²)			Values of bulk Modulii B (10 ¹⁰ N/m ²)	
	C ₁₁	C ₁₂	C ₄₄	Calc. (Present.)	Calc.(1.)
0.00	18.885	9.334	6.373	4.321	
0.50	17.732	9.363	6.125	4.221	
1.00	17.321	9.321	5.987	4.132	
1.50	16.421	9.233	5.635	4.110	
2.00	15.321	9.121	5.334	4.009	

Giri and Mitra 1986, which are in good agreement to the experimental values, have also applied this technique to KCl and KBr system for a number of compositions. Our theoretical study of binary liquid alloys at equi-concentration not only reproduces the concentration dependence of various elastic constants and microscopic properties but also throw light on the phenomenon of compound formation in its melt.

4 REFERENCES

[1] A. Lichanot, Solid State Communication 116 2000
 [2] Anit K. Giri And G.B. Mitra, Indian J. Phys. 59A, 318 1985
 [3] A. K. Giri And G.B. Mitra, Phys, stat, sol. (b) 134 K11. 1986
 [4] Bhatia, A.B. and Singh, R.N., Phys. Chem. Liq. 11 285 1982
 [5] C.Varotsos and M.Lazaridou, Phys. Stat.Sol.(b)129 K 95 1985
 [6] C.V. Pandya. P.R. Vyas, T. C. Pandya, Physica B 307 2001
 [7] Evens, R and Greenwood, D., Liquid Metals (Published by the institute of phys. London) 1976
 [8] Faber, T.E. and Ziman, J.M., Phil. Mag 11 153. 1965
 [9] Francisco J.Solis and Monica Olvera, Physical Review letter ;84–3350 2000
 [10] Harrison, W.A., Pseudopotentials in the Metals 1966

[11] Heine, V. and Abarenkov, I.V., Phil. Mag. 9 451. 1964
 [12] H.E. Karaca, I. Karaman, Scripta Mater 51 (3)367,20 2004
 [13] M. D. Segall, P.J.P, Condens Matter M 2002
 [14] N.Jha,A.K.Mishra and S.M.Rafique., Indian J.Phys 5A(5), 519–523 2001
 [15] P.Varotsos, Phys.Stat.Soi(b) 99, K. 93 1980
 [16] P.Varotsos and K. Alexopoulos, J.Phys.Chem. Solids 1985
 [17] P. Soderlind, J.A.Moriorty, Phy Rev B 57 10340 1998
 [18] R. Novokovic, D. Zivkovic, J. Mater Sci 40 2005
 [19] R.N. Singh ,F. Sommer, Phys Chem. Liquid 36 1998
 [20] B.B. Sahay and Srivastava, P.L., Canad. J. Phys. 58 401. 1980
 [21] R.N. Singh, Can. J. Phys. 65 309. 1987
 [22] R.N. Singh and Choudhary R.S., J.Phys.F.Metal Phys. 111577. 1981
 [23] R.N. Singh and Singh, S., Physica 128 B 304. 1985
 [24] T. Zhang,Anoue, Mater Sci Eng A 375–377 2004
 [25] T.T. Soma, H. Satoch Solid state Comm. 1981
 [26] L. Louail,D. Maouche, A. Roumili Materials Chemistry And Physics 2005
 [27] Young, W.H., Can. J. Phys 65 241 1987
 [28] Vashishta, P. and Singwi, K.S. Phys. Rev. B 6 875. 1972
 [29] Zhang, T.A Inoue Mater., Sci. Eng A 375,377 2000

Programmer Ranker Algorithm (PRA) for Evaluating Programmer Effort in the Context of Pair Programming

Manisha Giri, Meeta Dewangan

Abstract— Pair programming is a style of programming in which two programmers work side-by-side at one computer, continuously collaborating on the same design, algorithm, code, or test. In industry, the practice of pair programming has been shown to improve product quality, improve team spirit, aid in knowledge management, and reduce product risk. In education, pair programming also improves student morale, helps students to be more successful, and improves student retention in an information technology major. Project efficiency of pairs in program design tasks is identified by using pair programming concept. Pair programming involves two developers simultaneously collaborating with each other on the same programming task to design and code a solution. Programming aptitude tests (PATs) have been shown to correlate with programming performance. In this paper we will measure time productivity using pair programming, in two important ways: One is elapsed time to complete the task and the other is the total effort/time of the programmers completing the task. Using Programmer Ranker Algorithm (PRA) we will generate pair and Rank will be provided to each pair of Junior, Senior of industry. After providing rank the best pair is allocated to Embedded Software project type, Semi detached Software project type and Organic Software project type respectively.

Index Terms: Pair programming, PAT, Collaborative programming, Team building, PRA.



1. INTRODUCTION

Software applications grow larger and more complicated; these applications are then used in an infinite myriad of user systems. Perhaps, then, it is best for the complexity of these applications to be tackled by two humans at a time. The idea of pair-programming, two programmers working collaboratively on the same design, algorithm, code, or test, has independently emerged several times over the last decade. The practice of pair-programming is gaining popularity, primarily with the rise in the extreme Programming methodology [12]. The concepts underlying Pair Programming (PP) are not new [21], but PP itself has only recently attracted significant attention and interest within the software industry and academics.

Pair programming is a software practice that involves a pair of programmers simultaneously collaborating with each other on the same programming effort [12], [9], [16]. One programmer controls the keyboard and implements the program. The other programmer watches, identifies defects, and considers the direction of the work. Sitting side

by side at one computer, two colleagues collaborate on solving the problem, designing the algorithm, and coding.

Pairs regularly switch the driver and navigator roles and rotate their partners with other teams: This practice is thought to facilitate skills transfer and job rotation [15].

Some take the view that pair Programming is neither as economical nor as productive as individual programming [4], [7]. Others argue that more studies of pair programming productivity are needed [9], [7], [1], [18]. Some further explore pair programming such as side-by-side programming [8] and a mixed software practice of pair programming and individual programming [13], while others propose more traditional alternatives to pair programming such as reviews [17] and mutual programming [5], [4].

Several previous controlled experiments have validated the following quantitative benefits of pair programming over individual programming.

1. Significant improvements in functional correctness.
2. Various other measures of quality of the programs being developed.
3. Reduced duration (a measure of time to market), with only minor additional overhead in terms of total programmer hours (a measure of cost or effort)
4. Reduced the elapsed time and produced better software quality.

-
- *Manisha Giri is currently pursuing masters degree program in computer science and engineering from Chhatrapati Shivaji Institute of Technology, Durg (C.G.), India E-mail: manishagiri1@gmail.com*
 - *Meeta Dewangan is working as an assistant professor in Department of Computer Science Engineering Chhatrapati Shivaji Institute of Technology, Durg India, E-mail:meeta.dewangan@csitdurg.in*

One exception is an experiment that showed no positive effects of PP with respect to time taken no improved functional correctness of the software product compared with individual development [7], which essentially doubled the cost of development. However, the results of that experiment also suggested that the standard deviation of the development times and program sizes of the PP group was lower, suggesting that PP might be more predictable than individual programming.

Therefore, in controlled experiments where design related tasks were intermingled with coding, elapsed time was less and the quality was better for pairs. Another unique aspect of PP included the first ever assessment of the moderating effects of system complexity and programmer expertise.

Earlier studies reported that pair programming took more man hours than individual programming. One limitation of the previous experiment that was addressed to some degree in this experiment was that the task was much more complex/novel for the subjects.

In this project we will measure time productivity using pair programming, in two important ways: One is lapsed time to complete the task and the other is the total effort/time of the programmers completing the task. Using Programmer Ranker Algorithm (PRA) we will generate pair and Rank will be provided to each pair of Junior, Senior of industry. After providing rank the best pair is allocated to Embedded Software project type, Semi detached Software project type and Organic Software project type respectively.

The remainder of this paper is organized as follows: Section 2 provides a brief history of the use of pair programming. Section 3 identifies the problem in the existing system. Section 4 explains our approach of pair programming that is Programmer Ranker Algorithm(PRA). Section 5 provides the pair programming results. The final section provides concluding remarks and points some possible directions for future research.

2. BACKGROUND

Since as early as 1991, cognitive researchers have been interested in how two programmers collaborate on the same task [19]. They reported that two programmers in a pair could generate more diverse plans and explore a larger number of alternatives than an individual programmer. In a faithful reenactment of a pair programming episode by two pair programming practitioners reported in [20], the pair spent more time talking, casually reasoning about requirements realization, data modeling, data structures, and semantic analysis, than discussing lexical analysis, syntax analysis, libraries of a computer language, and the integrated development environment (IDE). This suggests that pair programming may have benefits in situations such as design-related tasks, where alternative exploration can improve the solution. Some studies on pair programming

attempted to simulate complex real situations so as to provide a rich picture of the behaviors inherent in pair programming. However, these programming tasks could not be done at a single time and the experimental task had to be split [16], [6]. The length of time between two experimental sessions can variably affect results [16], [6]. Oftentimes, like the real world requirements, descriptive programming tasks, instead of symbolic ones, were given to participants. As a result, some individuals misunderstood the problem, even at the beginning of the experiment [16], [6]. These studies have provided valuable information related to pair programming. However, because they were not strictly controlled experiments, it is difficult to sort out influences on the results.

3. PROBLEM IDENTIFICATION

Programming teams [3] in industry in which pair programming was practices report significantly improved team work among the members. If the pair can work together, then they learn ways to communicate more easily and they communicate more often. In many cases, these industrial teams continually rotate partners; two people do not work together for more than a short increment. This increases the overall information flow and team jelling farther.

Analysis related to a multivariate model[23] that expresses Individual Performance as a predictor of Pair Performance. Moreover, Personality was included as a predictor of Individual Performance as shown in fig 1.

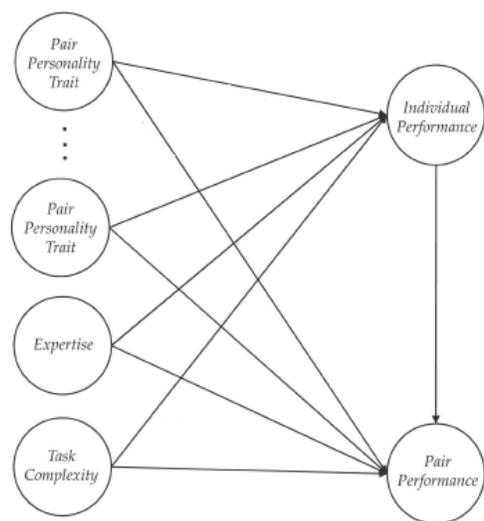


Fig. 1. Multivariate conceptual model(Individual Performance vs Pair Performance).

Earlier studies reported that pair programming took more man hours than individual programming. One limitation of the previous experiment that was addressed to

some degree in this experiment was that the task was much more complex/novel for the subjects. As we know there are three types of Software projects, Embedded Software project type, Semi detached Software project type and Organic Software project required different knowledge and skill set so Using Programmer Ranker Algorithm (PRA) we will generate pair and Rank will be provided to each pair of Junior, Senior of industry. After providing rank the best pair is allocated to Embedded Software project type, Semi detached Software project type and Organic Software project type respectively.

4. OUR APPROACH

In pair programming, time productivity [24] can be measured in two important ways: One is elapsed time to complete the task and the other is the total effort/time of the programmers completing the task.

The effort equation is as follows :-

$$E = a_b * (KLOC) b_b$$

$$D = c_b * (E) d_b$$

Where E - effort applied by per person per month,

D - Development time in consecutive months,

KLOC - estimated thousands of lines of code delivered for the project.

The coefficients a_b, c_b, and the coefficients b_b, d_b are given in the Table:

Table 1: Coefficients & exponents used in the Basic COCOMO Model

Software Project Type	a_b	b_b	c_b	d_b
Organic	2.4	1.1	2.5	0.4
Semi-detached	3	1.1	2.5	0.4
Embedded	3.6	1.2	2.5	0.3

From above table we can say that Embedded Software project type require more effort as compare to Semi-detached and Organic. Hence Embedded Software project should be allotted to high ranked pair.

4.1 PROGRAMMING APTITUDE TESTS(PATs)

Programming Aptitude Tests (PATs) can be correlated with programming performance. Aptitude tests in problem

solving and algorithm design can be used to test the effect of pairs in these tasks. The results of PATs can be used to generate pairs. PAT score will calibrate as follows:-

Time productivity can be measured in two important ways: One is elapsed time to complete the task and the other is the total effort/time of the programmers completing the task. Both important measurements of time can be incorporated in a single measurement, that is, the Relative Effort Afforded by Pairs (REAP)[14]:-

$$REAP = \left\{ \frac{\text{finish_time_of_pair} \times 2(\text{finish_time_of_individual})}{\text{Finish_time_of_individual}} \right\}$$

There are five cases to consider with REAP:

1. REAP < 0 When REAP is negative, the total time of pair programmers is less than the time of the individual programmer.

2. REAP ≈ 0 If REAP is zero, this is a break-even point, where the total time of pair programming is the same as individual programming.

3. REAP is between 0 and 100 When REAP is greater than zero but is less than 100 percent, pairs require more total man hours to complete the task but are faster than individual programmers.

4. REAP ≈ 100 If REAP is around 100 percent, the elapsed time for pair programmers is almost the same time as in the individual programmer.

5. REAP > 100 When REAP is greater than 100 percent, then the elapsed time for pair programming is longer than the time for an individual programmer.

4.2 PROGRAMMER RANKER ALGORITHM(PRA)

Procedure Gen_Pair()

//indiTime -> Finish Time of Individual

//p1Time -> Pair-I Finish Time

//p2Time -> Pair-II Finish Time

//p3Time = -> Pair-III Finish Time

REAP1 = (((p1Time * 2) - indiTime) / indiTime) * 100;

REAP2 = (((p2Time * 2) - indiTime) / indiTime) * 100;

REAP3 = (((p3Time * 2) - indiTime) / indiTime) * 100;

if (REAP1 < REAP2)

{

if (REAP1 < REAP3)

{

"The Pair One is Best compare to the others";

}

else

{

"The Pair three is Best compare to the others";

```

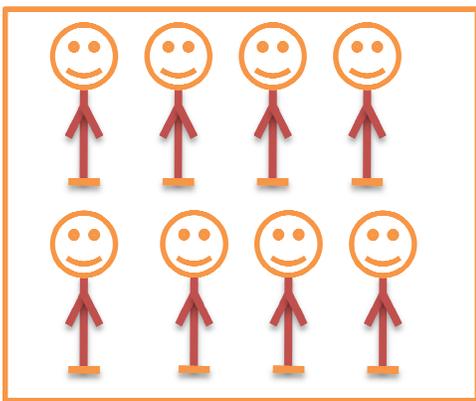
    }
  }
  else if (REAP1 > REAP2)
  {
    if (REAP2 < REAP3)
    {
      "The Pair two is Best compare to the others";
    }
    else
    {
      "The Pair three is Best compare to the others";
    }
  }
}

```

End Gen_Pair

Pair will be generated among Junior and Senior Staffs of industry. Now our next procedure will evaluate correct pair among different pairs generated using Gen_Pair procedure.

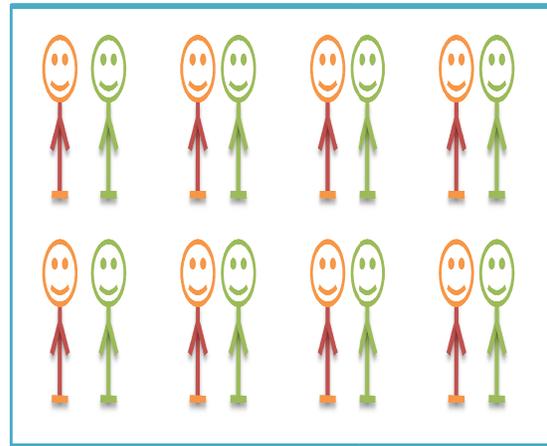
Junior Staff (Individual)



Senior Staff (Individual)



Pair of Junior and Senior Staff of Industry(Pairs)



Procedure Pair_Rank()

```

// p1Time -> Pair-I Finish Time
// p2Time -> Pair-II Finish Time
// p3Time -> Pair-III Finish Time
.
.
.
.
// pnTime -> Pair-n Finish Time

//T=          /n
//n -> Total Number of Pair

for(i=1; i<=n ; i++)
{
  T=T+piTime
}
T=T/n
//pi[n] -> To Store piTime

//Sorting pi
SORT_PAIR(A, p, r)
  if p < r
  then q ← PARTITION(A, p, r)
  SORT_PAIR(A, p, q - 1)
  SORT_PAIR(A, q + 1, r)
End SORT_PAIR

PARTITION(A, p, r)
  x ← A[r]
  i ← p - 1
  for j ← p to r - 1
  do if A[j] ≤ x
  then i ← i + 1
  exchange A[i] ↔ A[j]
  exchange A[i + 1] ↔ A[r]

  return i + 1
end PARTITION
End Pair_Rank

```

5. PAIR PROGRAMMING RESULTS

Anecdotal and empirical evidence reported in the literature suggest several organizational and personal benefits of PP over individual programming, such as reduced time to market, reduced development costs, improved quality of the software, reduced costs of training new personnel, and enhanced trust, motivation, and information and knowledge transfer among developers.

Fig 2 provides the comparison of pair programmers and individuals. It shows that the effort spent to develop the project can be reduced by pair programming. Programmer Ranker Algorithm (PRA) is used to generate pairs and the pairs generated by PRA can significantly reduce the Project development time and cost.

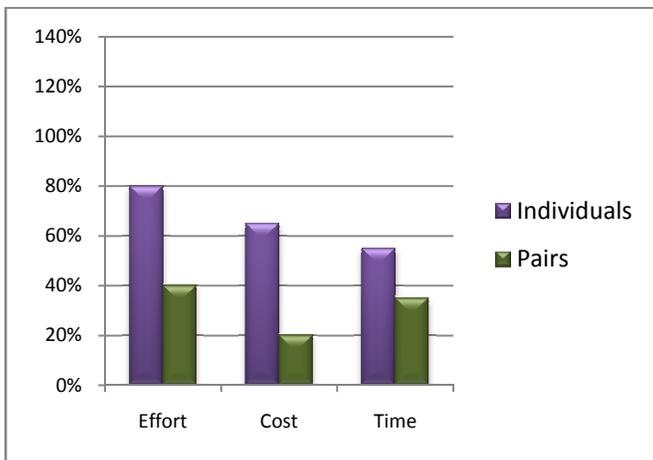


Fig 2. Comparison of pair programmers and individuals

6. CONCLUSION

The primary contribution of this study is to provide an overview of Pair Programming and to demonstrate the use of Programming Aptitude Test in the aspect of pair generation or team building that facilitates to make pair of newly hired programmers in an industry.

In our study, we have pointed out the use of PAT as a measurement of productivity and to evaluate the performance of individuals and pairs in order to generate the correct pairs. Our study showed that junior individuals may lack the necessary skills to perform tasks with acceptable quality, in particular, on more complex systems. Junior pair programmers achieved a significant increase in correctness compared with the individuals and achieved approximately the same degree of correctness as senior individuals. Software testing is often viewed as requiring less skill than initial system development and is thus often allocated to the more junior staff. Our study concludes that,

if juniors are assigned to complex tasks, they should perform the tasks in pairs.

Programmer Ranker Algorithm (PRA) will generate pair and Rank will be provided to each pair of Junior, Senior of industry. After providing rank the best pair is allocated to Embedded Software project type, Semi detached Software project type and Organic Software project type respectively. This will reduce the time and effort requires developing the Embedded Software project which will eventually reduce overall cost of software.

REFERENCES

- [1] A. Parrish, R. Smith, D. Hale, and J. Hale, "A Field Study of Developer Pairs: Productivity Impacts and Implications," *IEEE Software*, vol. 21, no. 5, pp. 76-79, Sept./Oct. 2004.
- [2] D.B. Mayer and A.W. Stalaker, "Selection and Evaluation of Computer Personnel: The Research History of SIG/CPR," *Proc. 23rd ACM Nat'l Conf.*, pp. 657-670, 1968.
- [3] F. P. J. Brooks, *The Mythical Man-Moth*: Addison-Wesley Publishing Company, 1975.
- [4] G. Keefer, "Extreme Programming Considered Harmful for Reliable Software," *Proc. Sixth Conf. Quality Eng. in Software Technology*, pp. 129-141, 2002.
- [5] G. Keefer, "Mutual Programming: A Practice to Improve Software Development Productivity," *Proc. Int'l Conf. Practical Software Quality and Testing '03*, 2003.
- [6] H. Hulkko and P. Abrahamsson, "A Multiple Case Study on the Impact of Pair Programming on Product Quality," *Proc. 27th Int'l Conf. Software Eng.*, pp. 495-504, 2005.
- [7] J. Nawrocki and A. Wojciechowski, "Experimental Evaluation of Pair Programming," *Proc. 12th European Software Control and Metrics Conf.*, pp. 269-276, Apr. 2001.
- [8] J. Nawrocki, M. Jasin`ski, L. Olek, and B. Lange, "Pair Programming versus Side-by-Side Programming," *Proc. 12th European Conf. Software Process Improvement*, pp. 28-38, Nov. 2005.
- [9] J. Nosek, "The Case for Collaborative Programming," *Comm. ACM*, vol. 41, no. 3, pp. 105-108, 1998.
- [10] J.M. Wolfe, "A New Look at Programming Aptitudes," *Business Automation*, vol. 17, pp. 36-45, 1970.
- [11] J.M. Wolfe, "Perspectives on Testing for Programming Aptitude," *Proc. 25th ACM/CSC-ER Ann. Conf.*, pp. 268-277, 1971.
- [12] K. Beck, *Extreme Programming Explained: Embrace Change*. Reading, Massachusetts: Addison-Wesley, 2000.
- [13] K.M. Lui and K.C.C. Chan, "Software Process Fusion: Uniting Pair Programming and Individual Programming Processes," *Proc. Int'l Software Process Workshop and Int'l Workshop Software Process Simulation and Modeling*, pp. 115-123, 2006.
- [14] Kim Man Lui, Keith C.C. Chan, and John Teofil Nosek "The Effect of Pairs in Program Design Tasks" *IEEE transactions on software engineering*, VOL. 34, NO. 2, march/april 2008
- [15] L. Williams and R.R. Kessler, *Pair Programming*

Illuminated. Addison-Wesley, 2003.

- [16] L. Williams, R.R. Kessler, W. Cunningham, and R. Jeffries, "Strengthening the Case for Pair Programming," *IEEE Software*, vol. 17, no. 4, pp. 19-25, July/Aug. 2000.
- [17] M. Ciolkowski and M. Schlemmer, "Experiences with a Case Study on Pair Programming," *Proc. First Int'l Workshop Empirical Studies in Software Eng.*, 2002.
- [18] M.M. Müller, "Two Controlled Experiments Concerning the Comparison of Pair Programming to Peer Review," *J. Systems and Software*, vol. 78, no. 2, pp. 166-179, 2005.
- [19] N. Flor and E. Hutcheins, "Analyzing Distributed Cognition in Software Teams: A Case Study of Team Programming During Perfective Software Maintenance," *Proc. Fourth Ann. Workshop Empirical Studies of Programmers*, 1991.
- [20] R.M. Martin, *Agile Software Development: Principles, Patterns and Process of Software Development*. Prentice Hall, 2001.
- [21] T.D. Cook and D.T. Campbell, *Quasi-Experimentation—Design & Analysis Issues for Field Settings*. Houghton Mifflin, 1979.
- [22] W.J. McNamara and J.L. Hughes, "A Review of Research on the Selection of Computer Programmers," *Personnel Psychology*, vol. 14, pp. 39-51, 1961.
- [23] Jo E. Hannay, Erik Arisholm, Harald Engvik, and Dag I.K. Sjøberg, "Effects of Personality on Pair Programming" *IEEE Transactions on software engineering*, vol. 36, no. 1, january/february 2010.
- [24] Nancy Merlo - Schett "COCOMO Model" Seminar on Software Cost Estimation Requirements Engineering Research Group Department of Computer Science University of Zurich, Switzerland, 2002 / 2003

LPCM Controller for Transformer less Dc–Dc Converter with High Step-Up Voltage Gain

P.Venkatesh, K. Ranganayakulu

ABSTRACT—Conventional dc-dc boost converters are unable to provide high step-up voltage gains due to the effect of power switches, rectifier diodes, and the equivalent series resistance of inductors and capacitors. In this paper a Linear Peak Current Mode Controller (LPCM) for a transformerless dc-dc converters is proposed to achieve high step-up voltage gain without an extremely high duty ratio. In the proposed converter, two inductors with same level of inductance are charged in parallel during the switch-on period and are discharged in series during the switch-off period. The structures of the proposed converter and controller are very simple. Only one power stage is used. Moreover, the steady-state analyses of voltage gains and boundary operating conditions are discussed in detail.

Index Terms—DC–DC boost converter, high step-up voltage gain, power stage.

1. INTRODUCTION

The DC-DC converter with high step up gain is used for many applications, such as high-intensity- discharge (HID) lamp ballast for automobile headlamps, fuel- cell energy conversion systems, solar-cell energy conversion systems and the battery back-up system for uninterruptible power supplies. Theoretically, the DC-DC boost converter can achieve high step-up voltage gain with an extremely high duty ratio. However, in practice, the step-up voltage gain is limited due to the effect of power switches, rectifier diodes and the equivalent series resistance (ESR) of inductors and capacitors. Also, the extremely high duty-ratio operation will result in serious reverse-recovery problem. Many topologies have been presented to provide high step-up voltage gain without an extremely high duty ratio. The DC-DC fly back converter is a very simple structure with high step-up voltage gain and electrical isolation, but the active switch of this converter will suffer high voltage stress due to the leakage inductance of the transformer. For recycling the energy of the leakage inductance and minimizing the voltage stress on the active switch, some energy-regeneration techniques have proposed to clamp the voltage stress on the active switch and to recycle the leakage-inductance energy. The coupled-inductor techniques provide solutions to achieve high voltage gain, low voltage stress on the active switch and high efficiency without the penalty of high duty ratio. The transformer less DC-DC converters, which include the cascade boost type, the quadratic boost type, the voltage-lift type, the capacitor-diode voltage multiplier type and the boost type integrating with switched-capacitor technique. However, these types are all complex and higher cost. For getting higher step-up voltage gain, the

other DC- DC converters are also presented. Compared with the converter as shown in fig.1, proposed converter has the following merits: (i) two power devices exist in the current-flow path during the switch-on period, and one power device exists in the current-flow path during the switch-off period. (ii) The voltage stresses on the active switches are less than the output voltage. (iii) Under the same operating conditions, including input voltage, output voltage, and output power, the current stress on the active switch during the switch-on period equals a half of the current stress on the active switch of the converter in fig.1. The proposed DC-DC converters fig.2 utilize the switched-inductor technique, which two inductors with same level of inductance are charged in parallel during the switch-on period and are discharged in series during the switch-off period, to achieve high step-up voltage gain without the extremely high duty ratio. The operating principles and steady-state analysis are discussed in the following sections. To analyze the steady-state characteristics of the proposed converters, some conditions are assumed as: (1)All components are ideal. The on-state resistance R_{DS} (ON) of the active switches, the forward voltage drop of the diodes and the ESRs of the inductors and capacitors are ignored.(2) All capacitors are sufficiently large and the voltages across the capacitors can be treated as constant. The modified boost type with switched-inductor technique is shown in fig.1.

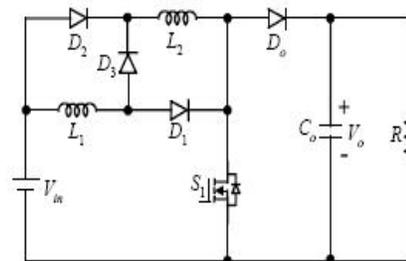


Figure 1 Tranformerless DC-DC high step-up converter

The structure of this converter is very simple. Only one power stage is used in this converter. However, this converter has two issues: (i) three power devices exist in the

- P.Venkatesh : is currently pursuing M.Tech scholar in EEE Dept., KL University, Green fields,Vaddeswaram Guntur Dt.-522502,Andhrapradesh,India,.pvenkatesh41@gmail.com
- K. Ranganayakulu: Assistant Professor in EEE Dept., KL University,Green fields, Vaddeswaram Guntur Dt., Andhrapradesh,India,. raga.kothamasu@gmail.com

current-flow path during the switch-on period and two power devices exist in the current-flow path during the switch-off period. (ii) The voltage stress on the active switch equals the output voltage. When switch S_1 is ON the output voltage is zero, during this period Inductors L_1 and L_2 are charged. When switch S_1 is OFF, the output voltage appears.

2. PROPOSED CONVERTER

A transformer less DC-DC high step-up converter is proposed as shown in figure 2. which consists of two active switches (S_1 and S_2), two inductors (L_1 and L_2) that have the same level of inductance, one output diode D_o , and one output capacitor C_o . Switches S_1 and S_2 are controlled simultaneously by using one control signal. Figure 2(d) and 2(e) shows some typical waveforms obtained during continuous conduction mode (CCM) and discontinuous conduction mode (DCM). The operating principles and steady-state analysis of CCM and DCM are presented in detail as follows.

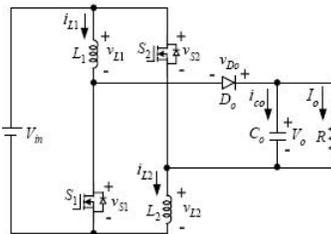


Figure 2 Proposed high step-up DC-DC converter I

2.1 CCM Operation:

The operating modes can be divided into two modes, defined as modes 1 and 2.

Mode 1 [t_0, t_1]:

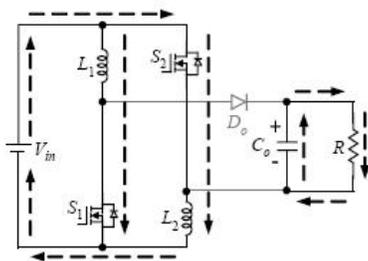


Figure 2 (a) Equivalent circuit of proposed converter I when switches ON

During this time interval, switches S_1 and S_2 are turned on. The equivalent circuit is shown in figure 2(a). Inductors L_1 and L_2 are charged in parallel from the DC Source and the energy stored in output capacitor C_o is released to the load. Thus, the voltages across L_1 and L_2 are given as:

$$V_{L1} = V_{L2} = V_{in} \quad (1)$$

Mode 2 [t_1, t_2]:

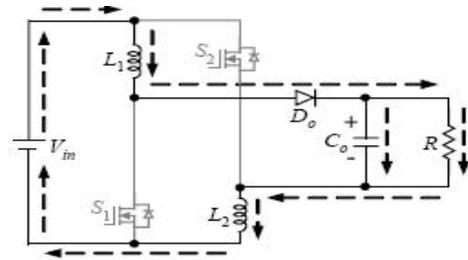


Figure 2(b) Equivalent circuit of proposed converter I when switches OFF

During this time interval, S_1 and S_2 are turned off. The equivalent circuit is shown in Fig. 2(b). The DC source, L_1 and L_2 are series-connected to transfer the energies to C_o and the load. Thus, the voltages across L_1 and L_2 are derived as

$$V_{L1} = V_{L2} = \left(\frac{V_{in} - V_o}{2} \right) \quad (2)$$

By using the volt-second balance principle on L_1 and L_2 , the following equation can be obtained:

$$\int_0^{DT_s} V_{in} dt + \int_{DT_s}^{T_s} \left(\frac{V_{in} - V_o}{2} \right) dt = 0 \quad (3)$$

Simplifying (3), the voltage gain is given by

$$M_{ccm} = \left(\frac{V_o}{V_{in}} \right) = \left(\frac{1 + D}{1 - D} \right) \quad (4)$$

From figure 2(d), the voltage stresses on S_1 , S_2 , and D_o are derived as

$$V_{s1} = V_{s2} = \left(\frac{V_o + V_{in}}{2} \right) \quad (5)$$

$$V_{D_o} = V_o + V_{in}$$

2.2 DCM Operation:

The operating modes can be divided into three modes, defined as modes 1, 2, and 3.

Mode 1 [t_0, t_1]: During this time interval, S_1 and S_2 are turned on. The equivalent circuit is shown in fig. 2(a). The two peak currents of L_1 and L_2 can be found as

$$I_{L1p} = I_{L2p} = \frac{V_{in}}{2} D T_s \quad (6)$$

Where L is the inductance of L_1 and L_2 .

Mode 2 [t_1, t_2]: During this time interval, S_1 and S_2 are turned off. The equivalent circuit is shown in Figure 2(b). The DC source, L_1 , and L_2 are series-connected to transfer the energies to C_o and the load. Inductor currents i_{L1} and i_{L2} are decreased to zero at $t = t_2$. Another expression of I_{L1p}

$$I_{L1p} = I_{L2p} = \left(\frac{V_o - V_{in}}{2} \right) D T_s$$

and I_{L2p} is given as

(7) Mode 3 [t₂, t₃]: During this time interval, S₁ and S₂ are still turned off. The equivalent circuit is shown in Fig. 2(c). The energies stored in L₁ and L₂ are zero. Thus, only the energy stored in C₀ is discharged to the load.

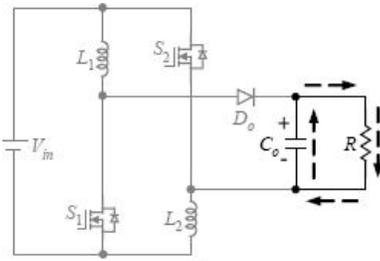


Figure 2(c) Equivalent circuit of proposed converter I when switches OFF in DCM mode

From (6) and (7), D₂ is derived as follows:

$$D_2 = \left(\frac{2Dv_{in}}{v_o - v_{in}} \right) \quad (8)$$

From Figure 2(e), the average value of output-capacitor current during each switching period is given by

$$I_{co} = \left(\frac{\frac{1}{2}DT_2 I_{Lr} - I_o T_s}{T_s} \right) = \frac{1}{2} D I_{Lr} - I_o \quad (9)$$

Substituting (6) and (8) into (9), I_{co} is derived as

$$I_{co} = \left(\frac{D^2 v_{in} T_s}{L(v_o - v_{in})} \right) - \left(\frac{v_o}{R} \right) \quad (10)$$

Since I_{co} equals zero under steady state, equation (10) can be re-written as follows

$$\left(\frac{D^2 v_{in} T_s}{L(v_o - v_{in})} \right) = \left(\frac{v_o}{R} \right) \quad (11)$$

Then, the normalized inductor time constant is defined as

$$\Gamma_L = \left(\frac{L f_s}{R} \right) \quad (12)$$

where f_s is the switching frequency (f_s = 1/T_s).

Substituting (12) into (11), the voltage gain is given by

$$M_{DCM} = \left(\frac{v_o}{v_{in}} \right) = \frac{1}{2} + \sqrt{\left(\frac{1}{4} + \frac{D^2}{\Gamma_L} \right)} \quad (13)$$

Typical waveforms for proposed converter I,

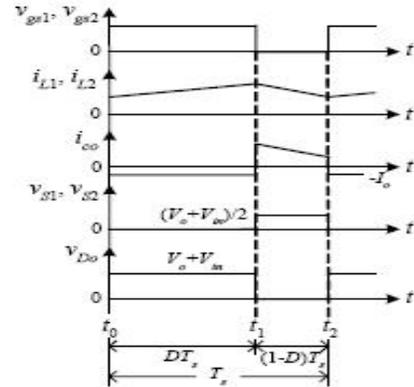


Figure 2 (d) CCM operation

2.3 Boundary Operating Condition between CCM and DCM

If proposed converter I is operated in boundary conduction mode (BCM), the voltage gain of CCM operation equals the voltage gain of DCM operation. From (4) and (13), the boundary normalized inductor time constant τ_{LB} can be derived as follows:

$$\Gamma_{LB} = \frac{D(1-D)^2}{2(1+D)} \quad (14)$$

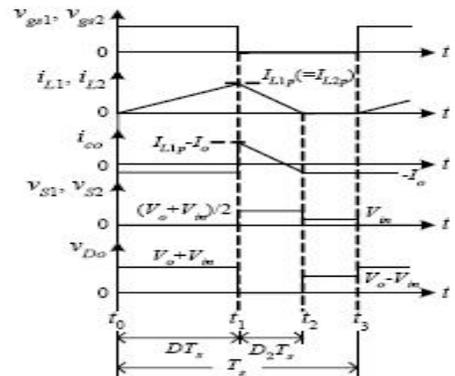


Figure 2(e) DCM operation.

The curve of τ_{LB} is plotted in Figure 2(f). If τ_L is larger than τ_{LB}, proposed converter I is operated in CCM

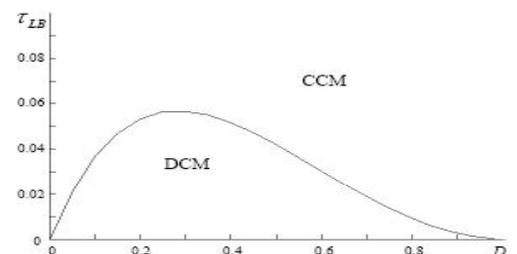


Figure 2(f) Boundary condition of proposed converter I

3. LINEAR PEAK CURRENT MODE CONTROL

Linear Peak Current Mode Control (LPCMC)-enables CCM operated rectifiers to be controlled using a much simpler controller Fig.3 (a).LPCMC offers the following advantages: Elimination of the controller multiplier and input voltages sensing circuits, unconditional stability of the current loop, and ease of implementation using low standard PWM control IC's.The control technique is based on designing a current loop whose static gain is linearly dependent upon the off-duty cycle of the switch

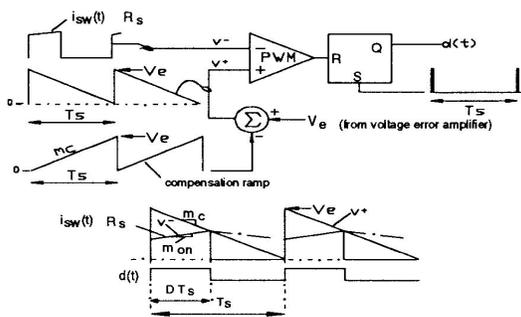


Figure 3(a) Linear Peak Current Mode Controller

The peak inductor current is expressed as

$$i_L(\text{peak}) = \frac{V_e}{R_s} \tag{15}$$

Where $I_{ref} = \frac{V_e}{T_s}$

V_e is the voltage error amplifier output signal, and R_s is the current sensing resistor, the peak inductor current becomes

$$i_L(\text{peak}) = I_{ref} \frac{m_c T_s}{R_s} D \tag{16}$$

m_c is the slope of the compensating ramp, and T_s is the switching period, and D is the duty cycle. Consider re-writing equation (16) in terms of $D'=1-D$

$$i_L(\text{peak}) = I_{ref} - \frac{m_c T_s}{R_s} D' \tag{17}$$

By rearranging equation (17),we can express the static gain of current loop as

$$i_L(\text{peak}) = \left(\frac{V_e}{R_s} - \frac{m_c T_s}{R_s} \right) + \frac{m_c T_s}{R_s} D' \tag{18}$$

Equation (18) shows a positive dependence of the static current loop gain of the off-duty cycle D' .Figure 3.2 plots $i_L(\text{peak})/I_{ref}$ versus D' .

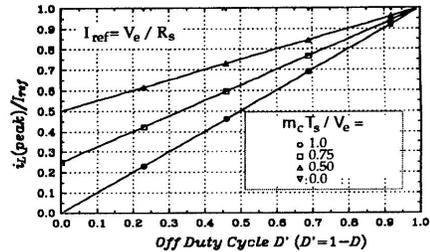


Figure 3(b) Reference current to peak current static gain
From equation (18), it is apparent that by choosing the compensation ramp appropriately, the first term will cancel, and the peak inductor current will be linearly related to D' .

$$i_L(\text{peak}) = \frac{m_c T_s}{R_s} D' \tag{19}$$

Where $\frac{m_c T_s}{V_e} = 1$, or $m_c = \frac{V_e}{T_s}$ (20), (21)

With the compensating slope defined by equation (21), equation (19) can be rewritten as

$$i_L(\text{peak}) = \frac{V_e}{R_s} D' \tag{23}$$

Equation (19) and (23) both reveal the linear relationship between the peak inductor current and the off-duty cycle D' .Hence the name Linear peak current control.

4. SIMULATION RESULTS

A Linear peak current mode control (LPCM) transformer less dc-dc converters with high step-up voltage gain has been simulated using MATLAB/Simulink. The simulation diagram of proposed controller to is shown in figure4.Simulation results are shown in fig.5 and fig.6.

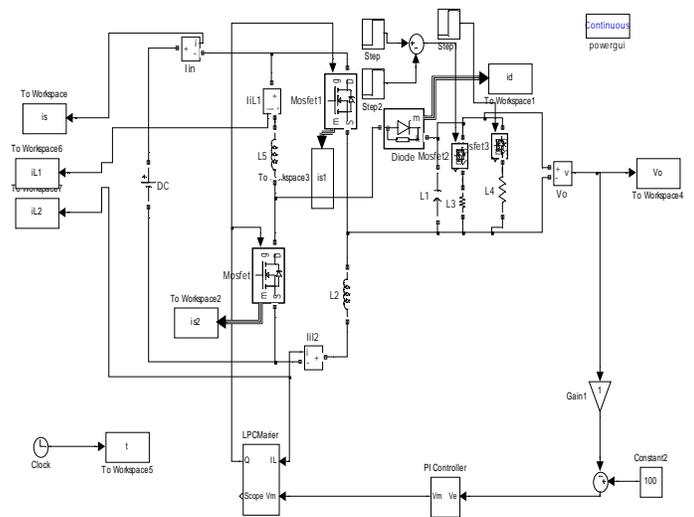


Fig4. Simulation diagram of LPCM Controller For Transformer Less Dc-Dc Converter With High Step-Up Voltage Gain

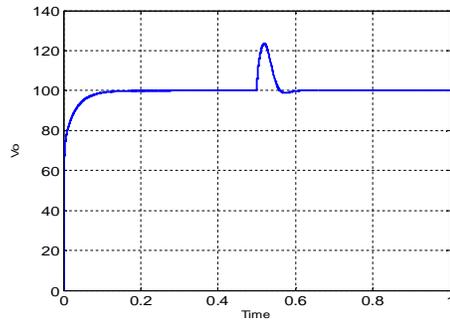


Fig5. Output voltage when load is changes from 40W to 5W at t=0.5Sec.

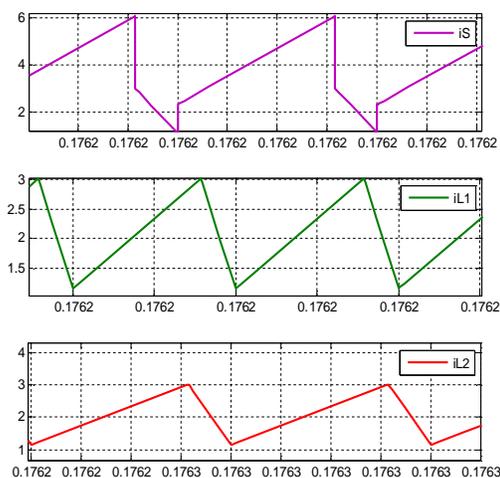


Fig6. Switch current, Inductor1 Current, Inductor2 Current waveform of Transformer Less Dc–Dc Converter With High Step-Up Voltage Gain.

5. CONCLUSION

A linear peak current mode control technique has been presented which enables simple, low cost. The technique possesses an inherently stable current loop, and the outer voltage loop is designed in a fashion similar to the other current mode control techniques. This paper has studied LPCM controller for transformerless dc–dc converter with high step-up voltage gain. Since the voltage stresses on the active switches are low, active switches with low voltage ratings and low ON-state resistance levels $R_{DS(ON)}$ can be selected. The steady-state analyses of the voltage gain and the boundary operating condition are discussed in detail. Finally the controller concept was generalized to include average current mode controller. In this case the sensed signal was a filtered version of the inductor current. In LPCM controller multiplier present in conventional current mode controller is eliminated as a result of making profitable use of the inherent duty cycle dependent modulator gain.

6. REFERENCES

- [1.] D. C. Lu, K. W. Cheng, and Y. S. Lee, "A single-switch continuous conduction- mode boost converter with reduced reverse-recovery and switching losses," *IEEE Trans. Ind. Electron.*, vol. 50, no. 4, pp. 767–776, Aug. 2003.
- [2.] B. R. Lin and F. Y. Hsieh, "Soft-switching zeta-fly back converter with a buck-boost type of active clamp," *IEEE Trans. Ind. Electron.*, vol. 54, no. 5, pp. 2813–2822, Oct. 2007.
- [3.] T. F. Wu, Y. S. Lai, J. C. Hung, and Y. M. Chen, "Boost converter with coupled inductors and buck-boost type of active clamp," *IEEE Trans. Ind. Electron.*, vol. 55, no. 1, pp. 154–162, Jan. 2008.
- [4.] R. J. Wai and R. Y. Duan, "High-efficiency DC/DC converter with high voltage gain," *Proc. Inst. Elect. Eng.—Elect. Power Appl.*, vol. 152, no. 4, pp. 793–802, Jul. 2005.
- [5.] B. Axelrod, Y. Berkovich, and A. Ioinovici, "Transformerless DC–DC converters with a very high DC line-to-load voltage ratio," in *Proc. IEEE ISCAS*, 2003, pp. III-435–III-438.
- [6.] O. Abutbul, A. Gherlitz, Y. Berkovich, and A. Ioinovici, "Step-up switching-mode converter with high voltage gain using a switched capacitor circuit," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 50, no. 8, pp. 1098–1102, Aug. 2003.

Real Time Patient Tele-monitoring System Using LabVIEW

Mr. Bhavin Mehta, Ms. Divya Rengarajan, Mr. Ankit Prasad

Abstract—Patient tele-monitoring is remotely monitoring the vital parameters of patient and providing them to the doctor at remote location than the patient thus ensuring mobility of both patient and the doctor. In the present paper the physiological parameters such as Electrocardiogram, Heart rate, SPO₂, Phonocardiogram and Temperature are obtained processed and displayed in a graphical user interface, then provided dynamically to a web page in real time to be viewed by an authorized doctor, if anyone of the vital parameter go out of normal range than an alert file is generated by the system. Finally alerts will be given in twoways, firstly on the doctor's mobile about the vital signs of the patient using Android application and secondly alert given by an email notification. All the three objectives have been implemented using LabVIEW.

Index Terms—LabVIEW, Patient tele-monitoring, Electrocardiogram, Heart rate, SPO₂, Phonocardiogram, Temperature, Patient monitoring system (PMS), Web page, Mobile, Email.

1 INTRODUCTION

The modern visionary of healthcare industry is to provide better healthcare to people anytime and anywhere in the world in a more economic and patient friendly manner. Therefore for increasing the patient care efficacy, there arises a need to improve the patient monitoring devices and make them more mobile. The medical world today faces two basic problems when it comes to patient monitoring, firstly the need of healthcare providers present bedside the patient and secondly the patient is restricted to bed and wired to large machines. In order to achieve better quality patient care, the above cited problems have to be solved. As the bioinstrumentation, computers and telecommunications technologies are advancing, it has become feasible to design a home based vital sign telemonitoring system to acquire record, display and transmit the physiological signal from the human body to any location. [1] Remote patient telemonitoring system using Java enabled 3G mobile phone enables doctors to monitor the vital biosignal such as *ECG*, *Respiration*, *SPO₂* and so on of patients in ICU/CCU using the real time waveform and data monitoring function of installed Java based application on the mobile phone. [2] The need for real time notification of vital signs of patient to the doctor is of prime importance, thus the need of active database system arises, that is grouped with patient monitoring device. The JMS was used to form the notifications. [3] The importance of PMS in medical treatment is very high; therefore the medical manufacturers are introducing centralized PMS. In centralized PMS all patient monitors are connected with a single server based PMS. The TCP/IP protocol suite based architecture systems are capable to upgrade the

PMS's firmware and software through dedicated TCP/IP protocol suite via open communication network. For a more efficient patient care by the caregivers, the PMS must be incorporated with smart alarm systems. [4] Thus the new PMS with intelligent alarm system has significantly improved sensitivity and specificity of monitoring and also demonstrated the feasibility of real time learning at the bedside. In order to alert the patient caregivers about the vital signs of patient these alarm systems are characterized based on the signal characteristics such as for ECG, attributes like R-R interval, peak detection and for other parameters their normal range sets the alarm system. [5] The computer based signal acquisition; processing and analysis system using LabVIEW is used as peak detection tool in ECG and as a filtering tool for biomedical signals. [6] The computer-based patient education can help improve the patient's awareness and understanding of his or her disease, thus the efficacy of treatment can be increased. This paper discusses the aspects of acquisition of physiological parameters like ECG, PCG, Temperature, SPO₂ and Heart rate, pre-processing them and displaying them in a graphical user interface for being viewed by the doctor and also sending the clinically useful data on a personalized website and alert message is generated by the system finally alerts will be given in two ways, firstly on the doctor's mobile about the vital signs of the patient using Android application and secondly by an email notification.

2 SYSTEM REPRESENTATION

The Real time vital parameter transmitting system schematic is shown in Fig. 1 with both electronic hardware and software components. The electronic component covers two aspects. The first ensures the acquisition and transmission of the signal using Acquisition card DAQmx, the second receives the signals on the server side using LabVIEW application and also is responsible for sending alert sms about that measurement information on physician's mobile.

- Bhavin Mehta is currently pursuing masters degree program in biomedical engineering in SRM University, India, E-mail: bhavin.bme@gmail.com
- Divya Rengarajan is currently pursuing masters degree program in biomedical engineering in SRM University, India, E-mail: deevyaa@gmail.com
- Ankit Prasad is currently pursuing masters degree program in masters of computer applications in SRM University, India.

3 MATERIALS AND MEHTODS

This section discusses the basics of LabVIEW, signal acquisition, processing and transmission.

3.1 LabVIEW

The LabVIEW software is used as the integrating platform for acquiring, processing and transmitting the physiological data as it is an excellent graphical programming environment to develop sophisticated measurement, test, and control systems using intuitive graphical icons and wires that resemble a flow-chart. The software also includes number of advanced mathematics blocks for functions such as integration, filter and other specialized capabilities. The LabVIEW Professional Development System allows creating stand-alone executables and the resultant executable can be distributed an unlimited number of times. The run-time engine and its libraries can be provided freely along with the executable.

3.2 Signal Acquisition

The signal acquired here are ECG, SPO₂ PCG, TEMEPERATURE .These signals acquired through sensors connected to PC having a LabVIEW platform through a DAQ device interfaced through an ELVIS Prototyping board having analog and digital input/output channels. The acquired signals are made available by the DAQ user interface in LabVIEW for further analysis that can be designed in the block diagram panel.

3.3 ECG Recording

The low noise ECG signal is acquired by National Instrument Educational Laboratory Virtual Instrumentation Suite (NI ELVIS) using 3-lead system. The acquired signal is further processed by LabVIEW which is having signal processing module. Signal processing tool kit contains digital filter and several transform such as Wavelet, Fast Fourier Transform (FFT) etc. The appropriate filter and transform have been implemented and this gives noise free signal and from that fast and reliable estimation of clinically important parameter such as R-R interval and ECG peaks accurately achieved using LabVIEW. In order to achieve the above functionalities, firstly baseline filter to removing the baseline noise and then signal applied to the wavelet denoise which performs noise reduction for 1D signal by using the discrete wavelet transform (DWT) and then passed the signal for Multiscale Peak Detection which used to detect peaks in a signal and from that each peak we determine the amplitude of each peak and width of signal.

3.4 Heart Rate Determination

Heart rate is determined by the number of heartbeats per unit of time, typically expressed as beats per minute (BPM). The measurement of heart rate is based on the Pan Tompkins algorithm for QRS detection. The algorithm is based on analysis of the slope, amplitude and width of QRS complexes. The algorithm includes a series of filters and methods that perform low pass, band pass, derivative, squaring, integration, adaptive thresholding and search procedure. Calculation of Heart rate based ECG graph collector present in LabVIEW that collects input signals and returns the most recent data, up to the speci-

fied maximum number of sample per channel. The maximum value of the peaks and pass them to waveform peak detection tool which defines the peak value for the particular given width and after that the peak value was divided by width and multiplied by 60 for the measuring heart rate in beats per minute, and then compared with normal heart rate range which gives the Bradycardia and Tachycardia conditions of the patient, when any of these conditions occurs sound and light alarms automatically start.

3.5 PCG Recording

The PCG (Phonocardiogram) is a vibration or sound signal related to the contractile activity of cardiohemic system and represent recording of heart sound signal. Recording of PCG requires a transducer to convert the vibration or sound signal into an electrical signal and then this signal was given to PCG amplifier. This setup was connected to the NI DAQ for acquiring the signal. After acquiring the signal, FFT was taken for getting the spectrum and then filtered by low pass filter to remove the noise form signal and thus giving accurate result of heart sound waveform.

3.6 SPO₂ Calculation

Pulse oximeter is a simple non-invasive method for monitoring the percentage of heamoglobin (Hb) saturated with oxygen (SPO₂).The pulse oximeter consists of probe attached to the patient's finger or ear lobe which is linked to NI ELVIS hardware for further processing. Pulse oximeter uses the basic principle of a pair of small LEDs operating at a different wavelengths; one red LED with a wavelength of 660nm, the other, an infrared LED with a wavelength of 910nm. The LEDs are designed to place opposite a photodiode that detects the transmitted light from LEDs. Absorption on each wavelength differs significantly for the oxyhaemoglobin and deoxyheamoglobin. Therefore from the difference of the absorption of the red and infrared light into ratio between oxy/deoxyhaemoglobin can be calculated and then this signal was given to pre-amplifier and low pass filter circuit for further processing. The electronic circuit for pulse oximeter is as shown in Fig 2. This set up was then connected to the DAQ for acquiring the signal in LabVIEW for calculating the ratio of absorption of oxyhaemoglobin and deoxyheamoglobin which was derived in LabVIEW by equation as in (1).

$$R = \frac{AC_R/DC_R}{AC_{IR}/DC_{IR}} \quad (1)$$

R= ratio of absorption of oxyhaemoglobin and deoxyheamoglobin.

The value of SPO₂was then calculated by the equation as in (2)

$$SPO_2 = (110 - 25R) \% \quad (2)$$

The value of SPO₂was then compared with the normal value which gives alert alarm, when the value of SPO₂decreases from the normal range. And that will be helpful to physician for further patient analysis.

3.7 Temperature Monitoring

The temperature sensing is performed by using a thermistor with excitation voltage of maximum +12v and GND based on the principle that the resistance changes with the temperature change. This is given by Steinhart-Hart equation as in (3).

$$1/T = a + b \ln(R) + c \ln^3(R) \quad (3)$$

$$\begin{aligned} \text{Where, } a &= 1.4 \times 10^{-3} \\ b &= 2.37 \times 10^{-4} \\ c &= 9.90 \times 10^{-8} \end{aligned}$$

This gives temperature in Kelvin, which was then converted into Celsius by the following formula as in (4).

$$\text{CELCIUS } [^{\circ}\text{C}] = [\text{K}] - 273.15 \quad (4)$$

Celsius was then converted into Fahrenheit by the following formula as in (5).

$$\text{FARENHEIT } [^{\circ}\text{F}] = [^{\circ}\text{C}] \times 9/5 + 32 \quad (5)$$

4 REAL TIME TELE-MONITORING USING LABVIEW

The LabVIEW has been used to build computer graphics interface (CGI) programs and URLs, to send and receive data using the Telnet protocol, to store and retrieve files from FTP servers and to publish VIs on the Web browser.

4.1 Web Publishing Tools

LabVIEW have internet toolkit including the G Web Server, which is an HTTP/1.0-compatible server used to run applications on the Web. Servers and CGI applications intercommunicate through environmental variables and standard inputs and output. When HTTP executes CGI VI, standard input data was generated as string and a request to send generate string as well is generated.

4.2 Publishing Front Panel with G Web Server

Here the G Web server was used to publish image of front panel on the Web. Using this static or animated front panel images can be loaded. The G Web server can generate images in JPEG or PNG image formats. All above described features have been used in this application to transmit data on Web. The acquisition is triggered by the physician remotely, once he decided to view the vital parameter of one of his patients, the request was redirected from the asp.net based website to G Web server which executed remotely a LabVIEW application embedded in the patient's pc and the session was opened, then the signal was displayed dynamically in the web page at the physician side. For this the patient is already prepared with the electrodes on, which means arrangement between patient and physician should have been prepared which is shown in Fig 3. Fig 4 shows all patient vital parameter signal accessed from the physician side. From this physician can retrieve all information about his patients that he may need while assessing their health condition, such as Heart rate, temperature, SPO₂ etc.

4.3 Alternative Recording Possibility

In some cases while assessing the health condition of a patient, need to go back to previous data occurs. Each session is saved in TDMS file now and then and then uploaded at the end of session to the server, saving this file the name and photo of the patient along with the session date, allows the physician to retrieve the desired session by date among all recorded session and patient information. Each and every process can be controlled by the physician; he can put a limitation on what recording time of signal to be saved.

4.4 Real Time Report Generation Using LabVIEW

Using report generation toolkit present in LabVIEW a real time patient record containing basic patient information such as name, age, gender and clinical information like temperature, spo₂, heart rate, ECG waveform and PCG waveform is generated.

4.5 Alerts on Mobile Using Android Application

Android is a software stack for mobile devices that includes an operating system, middleware and key applications. The SDK provides the tools and APIs necessary to begin developing applications on the Android platform using the Java programming. An application in MOTODEV Studio based on Eclipse was developed. An android 2.3.3 version phone for receiving alert message which is mainly developed using MOTODEV Studio was used. The code for this application is formed in java. Layout of the alert application is shown in Fig 5. This application runs through internet. When alert file is generated by the LabVIEW, it will be automatically uploaded using FTP on Web using AutoIt software. Alert application works continuously in background of mobile whenever it finds file, mobile vibrates and downloads file from the web and opens it in Microsoft office automatically. Requirement for operating application in physician mobile are android phone and internet connection in mobile.

4.6 Email Notification

This can be done using AutoIt software which works like scripting language designed for automating the Windows GUI and scripting. When alert file is generated by the LabVIEW, it will automatically send mail using Email notification application. Alert Email notification on mobile has been shown Fig.6.

5 RESULT AND DISCUSSION

This system can be used to transmit the patient vital parameter information in real time to remote location and can be viewed by the care giver as shown in Fig 4. Also a printable patient report can be generated any time as per the need. One such generated report was shown in Fig 7.

As this is medical application, reliability is needed in the first place. A reliable TCP protocol was used in this application which was implemented in LabVIEW. The system was tested for data acquisition for ten subjects and the following data was obtained. This has been represented in the interpretation table

as shown in table 1.

6 CONCLUSION

In this paper, telemonitoring application is presented which allows doctor to view his patient's vital parameter remotely and dynamically in a Web page in real time and does not need to have any special requirement on his PC; all he needs is an internet access.

For the patient side, it is a home based LabVIEW application embedded in a home PC, during signal acquisition.

7 FUTURE WORK

In future work the transport protocol (TCP/IP) can be replaced by the Real Time Transport protocol RTP, thus making it more secure, by implementing the RTP in the developed LabVIEW application.

ACKNOWLEDGMENT

The authors acknowledge Mrs. Angeline Kirubha and Dr. M. Anburajan of SRM University for their encouragement and guidance.

REFERENCES

- [1] Aleksandar C. Zoric, Sinisa S.Ilic, "PC Based Electrocardiography & Data Acquisition", TELSIKS, IEEE, pp 619-622, September 28- 30 2005.
- [2] Bergeron BMD. "Telemedicine in the practice setting: Infrastructure, not technology", limits practical application (editorial). Postgraduate Medicine 2003; Accessed February 23, 2004.
- [3] D .S. Benitez, A. zaidi, "virtual instrumentation for clinical assessment of cardiovascular and automatic functions" IEEE Proc- Sci. Measurement Technology, vol. 147, no. 6 pp 397-402, Medical Signal Processing, IEEE, Nov. 2000.
- [4] Eric P. Wildmaier, Hershel & Kevin T. Strang, 2007, "The Human Physiology,"Eleventh Edition, 11th edition. McGraw-Hill Higher Education
- [5] Gupta Sanjay, Joseph John 'Virtual instrumentation using LabVIEW", Electrical Engineering Series, Tata McGraw Hill pub. India, 2006.
- [6] Khandpur R. S, "A Handbook of Biomedical Instrumentation, Tata McGraw Hill pub; India, 1993.
- [7] Nguyen Quoc Cuong, Nguyen Thilan Huong, Pham Thi Ngoc Ven "Study and design of anElectrocardiograph" Proceedings of the first young Vietnamese Scientists Meeting (YVSM'05) Nha Trang, June 12-16, 2005
- [8] P. Ganesh Kumar, K. Thanushkodi,"A virtual instrumentation system for on line data acquisition in water treatment plant "WSEAS TRANSACTIONS on CIRCUITS and SYSTEMS Issue 6, Volume 5, June 2006
- [9] Raghu Korrapati, James A. Anderson, et.al "System modelling using virtual instruments" proceeding of IEEE southeast conference pp 121-126, 2002, southeast con.
- [10] Shuren Qin, zhong ji, hongjun zhu 'The ECG recording and analysis instrumentation based on virtual instrumentation technology and Continuous wavelet Transforms" proceeding of 25th international conference, IEEE, EMBS pp 3176-3179, September 17-21, 2003. Cancun, Mexico
- [11] A. Hernández, F. Mora, G. Villegas, G. Passariello, and G. Carrault, "Real-Time ECG Transmission via Internet for Nonclinical Applications," IEEE Tran INF Technol Biomed Vol. 1, No. 5, pp. 253-257, September 2001.
- [12] E. Zeki, Mehmet and Caglav, "Real-time ECG signal transmission via telephone network," Measurement, Volume 37, Issue 2, March 2005, Pages 167-171.
- [13] Patricia Mendoza, Perla Gonzalez, Brenda Villanueva, Emily Haltiwanger, and Homer Nazeran,, "A Web-based Vital Sign Telemonitor and Recorder for Telemedicine Applications," Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2004.
- [14] C. Wen, M. Yeh, K. Chang and R. Lee , "Real-time ECG telemonitoring system design with mobile phone platform", Measurement, Volume 41, Issue 4, May 2008, Pages 463-470
- [15] Carlos H. Salvador, Member, IEEE, M. Carrasco, M. González de Mingo, A. Carrero, J. Montes, L. Martín, M. Caverio, I. Lozano, and J. Monteagudo , "A GSM and Internet Services-Based System for Out-of-Hospital Follow-Up of Cardiac Patients," IEEE transactions on information technology in biomedicine, Vol. 9, no. 1, March 2005.
- [16] X. Ge, D. Lai, X. Wu, and Z. Fang, "A Realtime Continuous ECG Transmitting Method through GPRS with Low Power Consumption," China, second International Conference in Bio informatics and Biomedical Engineering, Shanghai, May 2008
- [17] J. Dong, S. Zhang and X. Jia, "A Portable Intelligent ECG Monitor Based on Wireless Internet and Embedded System Technology," International Conference on BioMedical. Engineering and Informatics. (BMEI 2008), Volume 2, Issue, 27-30 May 2008 pp. 553 – 556.
- [18] Hossein Fariborzi and Mahmoud Moghavvemi, "Wireless monitoring of cardiac activity: a critical review", International Journal of Biomedical Engineering and Technology, Volume 2, 2009, pp. 4-28
- [19] Schecke T, Langen M, Popp HJ, Rau G, Kasmacher H, Kalff G. Knowledge-based decision support for patient monitoring in cardioanesthesia. *International Journal of Clinical Monitoring and Computing*. 1992; 9(1): 1-11.
- [20] Siguira T, Mizushina S, Kimura M, Fukui Y, Harada Y. A fuzzy approach to the rate control in an artificial cardiacpacemaker regulated by respiratory rate and temperature: A preliminary report. *Journal of Medical Engineering and Technology*. 1991; 15: 107-110.

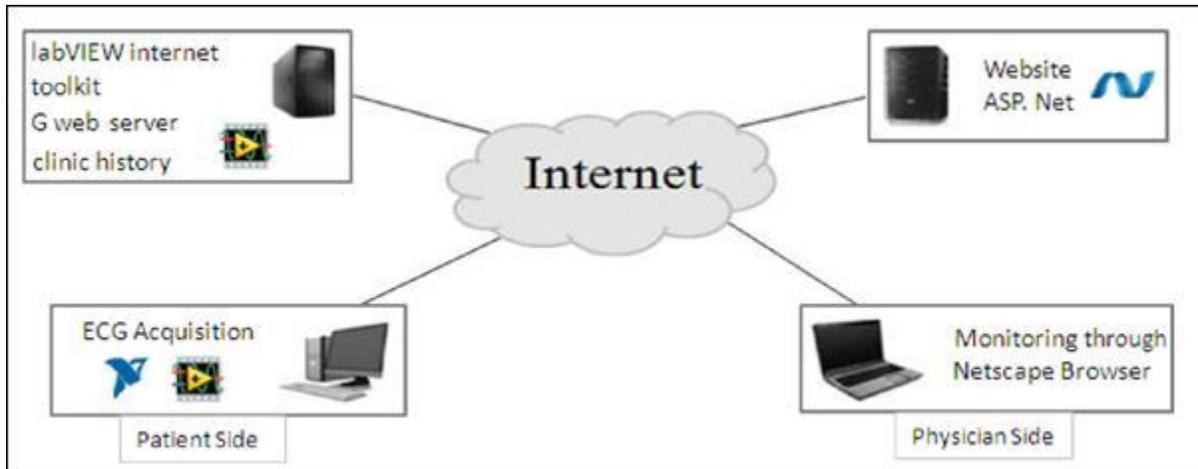


Fig. 1 Real Time Remote Monitoring Schematic

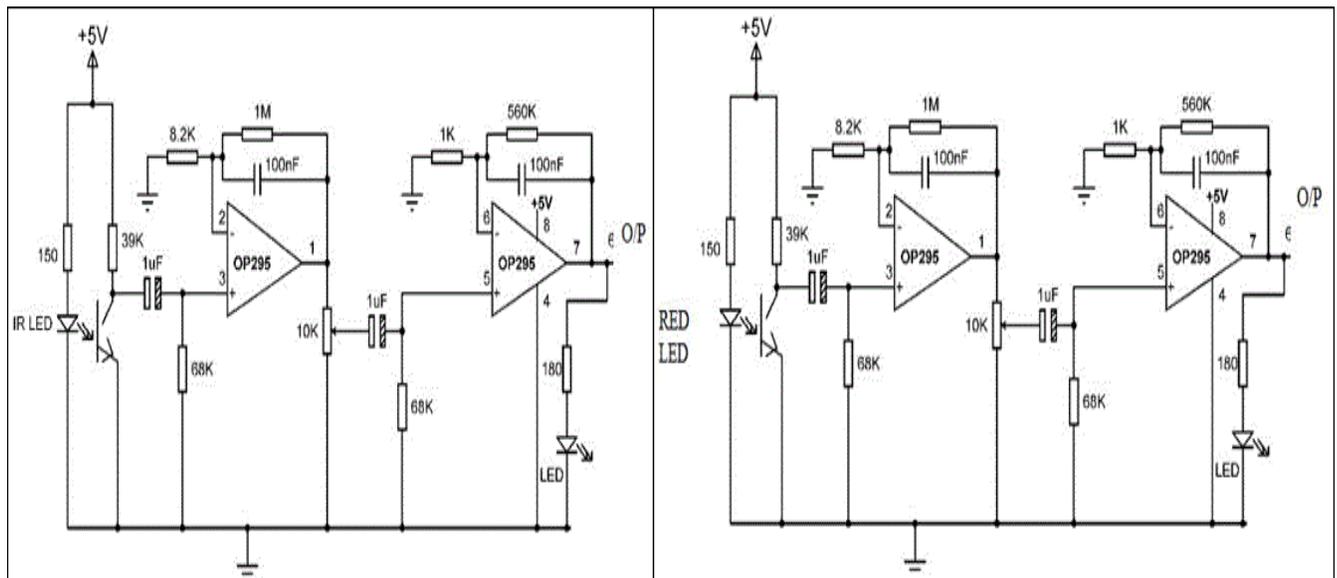


Fig. 2 Basic electronic circuit for pulse oximeter

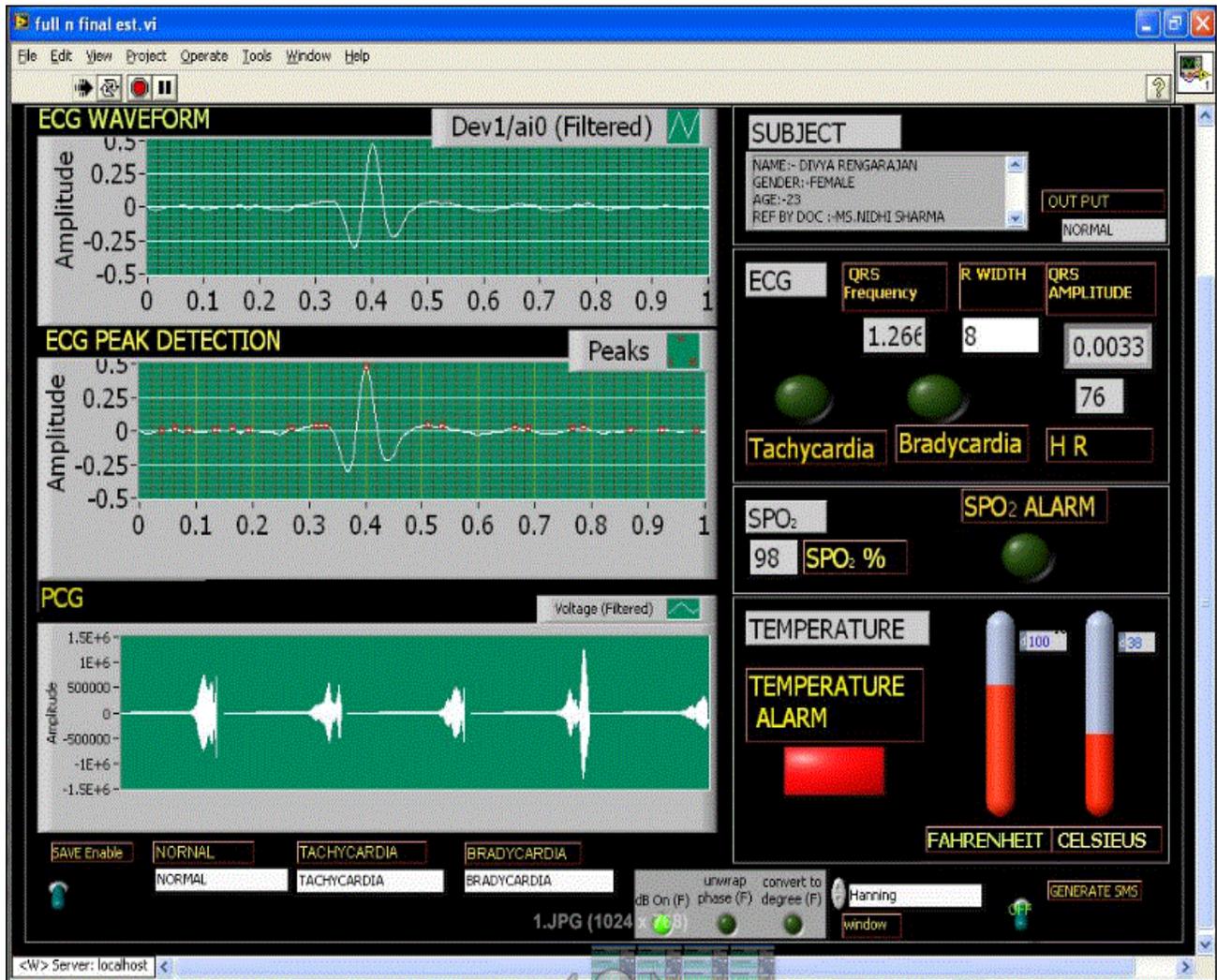


Fig. 3 Patient side monitoring

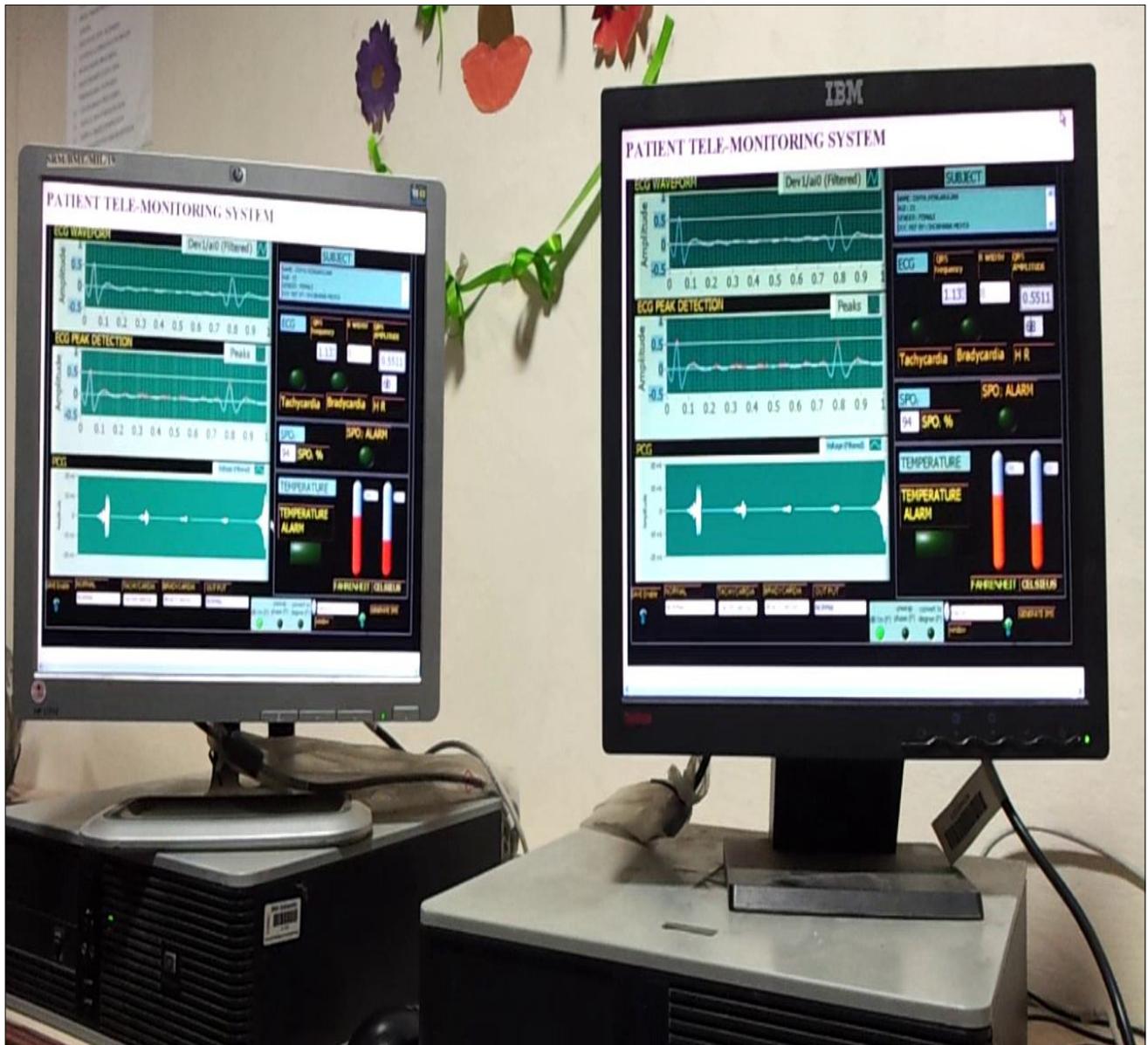


Fig. 4 Physician side, vital parameter of selected patient

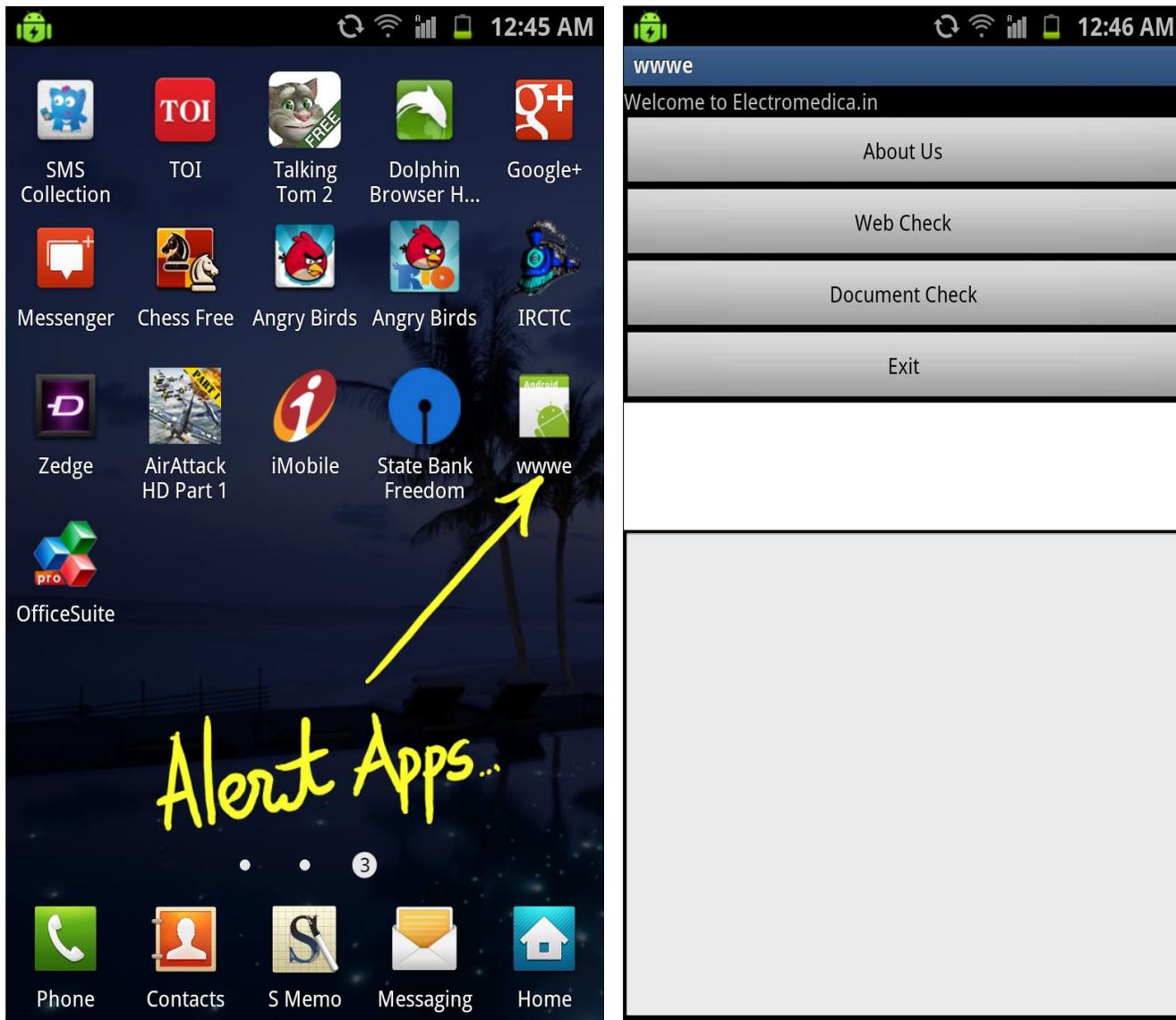


Fig.5 Alert application on Android mobile

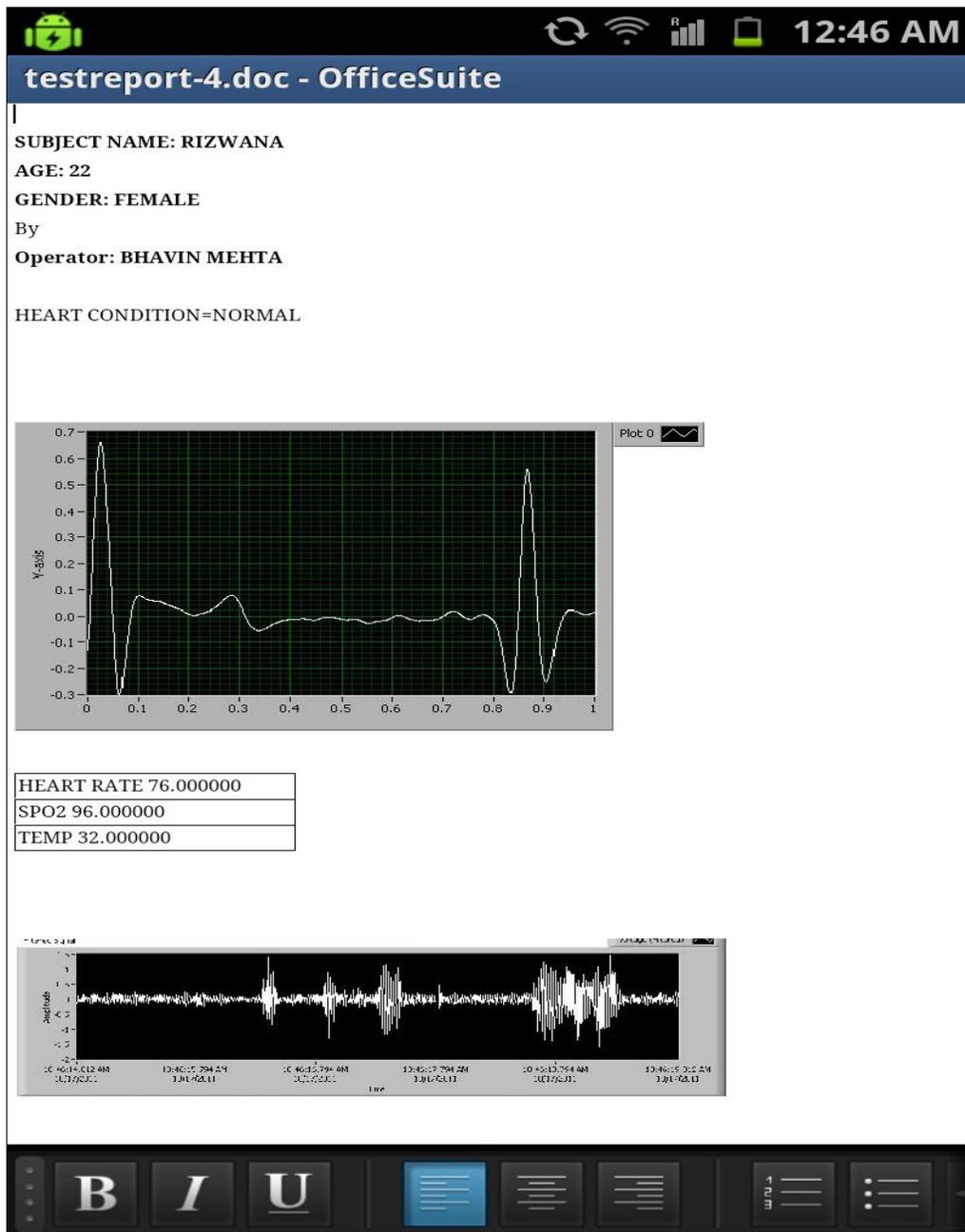


Fig.6 Alert Email notification on mobile

SUBJECT NAME: DIVYA

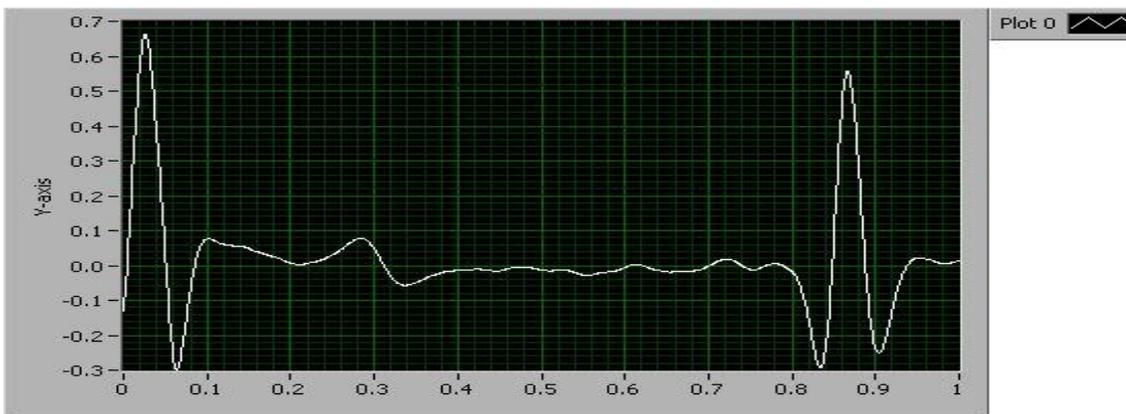
AGE: 22

GENDER: FEMALE

By

Operator: BHAVIN MEHTA

HEART CONDITION=NORMAL



HEART RATE 76.000000
SPO2 96.000000
TEMP 32.000000

PCG GRAPH

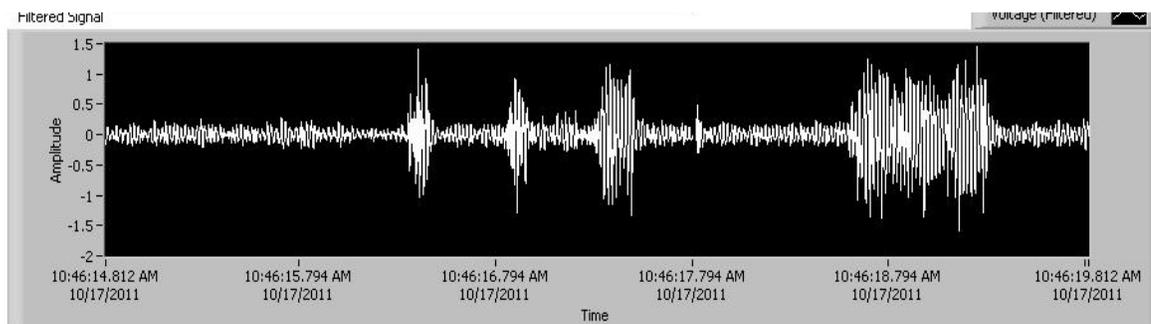


Fig.7 Real time printable patient report

TABLE 1
INTERPRTATION TABLE

CASE	TEMPERATURE (°C)	SPO ₂ (%)	HEART RATE (Beats/Min)	R-WAVE AMPLITUDE (mV)	QRS COMPLEX WIDTH (Sec)	INTERPRETATION
CASE 1	36	98	74	4.6	0.09	NORMAL
CASE 2	35	96	70	3.8	0.11	ABNORMAL
CASE 3	37	99	81	5.07	0.08	NORMAL
CASE 4	36	99	72	4.98	0.12	NORMAL
CASE 5	36	96	70	4.45	0.07	NORMAL
CASE 6	37	97	56	3.2	0.09	ABNORMAL
CASE 7	36	96	71	3.54	0.11	ABNORMAL
CASE 8	35	98	69	4.76	0.06	NORMAL
CASE 9	35	96	95	4.73	0.03	ABNORMAL
Case 10	37	99	90	4.52	0.12	NORMAL

NORMAL RANGE OF THE VARIOUS MEASURED PHYSIOLOGICAL PARAMETERS:

TEMPERATURE: (35-37) °C

SPO₂: 96% - 99%

HEART RATE (Beats/Min - bpm): 60-90bpm

R-WAVE AMPLITUDE (mV): >4.5 mV

QRS COMPLEX WIDTH (Sec): (0.04-0.12) mSec

Auditory Assessment Based On EEG

Ms. Divya Rengarajan, Mr. Bhavin Mehta, Ms. P. Vinupritha

Abstract— Auditory assessment is based on the determination of hearing threshold of the subject. The hearing threshold is minimum sound power level (dB) of the stimulus audible to subject. The subject's response to the stimulus is present in the electroencephalogram (EEG) as auditory evoked potential (AEP) once the stimulus is loud enough to excite the auditory system. The present paper reviews a real time auditory assessment method by plotting an audiogram on the basis of AEP using LabVIEW.

Index Terms—hearing threshold, auditory stimulus, auditory evoked potential (AEP), electroencephalogram (EEG), auditory loss, steady state auditory evoked potentials, audiogram, auditory steady state response (ASSR), amplitude modulated (AM), LabVIEW, virtual instrumentation (VI).

1 INTRODUCTION

Hearing impairment is a health disorder significantly seen in two ends of human life, approximately one in thousand new born infants and more than a quarter of adults over the age of 65 years have significant auditory loss. Early detection ensures provision of appropriate treatment at an early stage and thereby ensures development of normal speech and language in infants. In the case of adults this technique will help in choosing the perfect hearing aid. Conventional screening techniques use an audiometer which needs a behavioral response from the subject. In the case of elderly people and infants, they won't be able to provide any substantial response and in the other case which is of mentally challenged patients, they won't be willing to take the screening test. Thus this paper goes in the direction of reviewing another means to detect auditory loss so that early and accurate diagnosis is possible. This technique focuses on determining hearing thresholds at different frequencies based on the response of brain to the auditory stimulus through auditory evoked potential that are present in EEG. Auditory evoked potentials are small electrical potentials generated in brain in response to any auditory stimulus and recorded from scalp. The steady state auditory evoked potentials are the responses obtained when stimuli occurs at a rapid rate in such a way that the responses superimpose and cause a periodic responses at specific stimulation frequencies. This enables us to record the response at multiple frequencies and plot the frequency versus sound power level as an audiogram for further analysis. When the auditory evoked potentials are generated due to stimulus at a higher rate it gives rise to auditory steady state response (ASSR). Stimulus of amplitude modulated tones with carrier frequencies of 500, 1000, 2000, 4000Hz elicited ASSR and it was seen that there is no significant difference between single and multiple AM tones [1]. For the same stimuli the difference between mean evoked potential thresholds and behavioral pure

tone thresholds was taken and it was found that for 40Hz ASSR the difference was less compared to 80Hz ASSR [5]. The monotic ASSR is 0-10 dB higher than the behavioral ASSR threshold and also there is no difference between dichotic and monotic stimulus [2]. Chirp stimulus generates responses with larger amplitude than click stimulus and the responses at 40Hz repetition rate are larger than that at 80Hz [3]. The scalp distribution of both stimulus evoked and event related components of response vary with age yielding an equipotential distribution in older subjects. In children the latencies of event related potentials decreases with age [4]. It is also noted that the ASSR amplitudes and phases did not differ for single versus alternated stimulus polarities for both bone- and air-conduction stimuli [6]. Furthermore the AEP can provide the information about number and location of existing auditory nerve fibers as shown by using 40Hz AEP response [8]. This article discusses the method of stimulus generation, simultaneous EEG signal acquisition, analysis of the auditory evoked potential in the background EEG and plotting of an audiogram of the subject for the purpose of diagnosing him/her for hearing related disorder.

2 MATERIALS AND METHODS

This section discusses the hardware and software components used and also the process of signal acquisition, processing and analysis. The basic system representation is shown in fig.1. The system consists of both hardware and software components and can be used with a PC to collect and analyze auditory steady state responses and also facilitates research with ASSRs. This system can be also used as objective audiometry for hearing assessment. As the system should use standard hardware and flexible programming system, hence LabVIEW was used as a data acquisition platform.

2.1 System Requirements

The software and hardware components are as discussed below.

LabVIEW is a graphical programming language that lets us make programs in form of virtual instruments (VIs) which look and act like real instruments and also can be interfaced with many hardware devices using data acquisition cards. For

- Divya.Rengarajan is currently pursuing masters degree program in biomedical engineering in SRM University, India, E-mail: deevyaa@gmail.com
- Bhavin Mehta is currently pursuing masters degree program in biomedical engineering in SRM University, India, E-mail: bhavin.bme@gmail.com
- Ms. P. Vinupritha is currently working as an assistant professor (O.G) in deptt. Of Biomedical Engineering, SRM University, India.

the data acquisition board National Instrument's ELVIS Prototyping Board is used which has nearly seven physical channels, each of which is differential in nature. In this the signals are differentially amplified to a level that the signals usually in range of $10\mu\text{V}$ can be conveniently viewed in the PC.

DAQ hardware which is a basic A/D converter provides the interface between the real world signals and the PC containing LabVIEW. It can be in the form of modules that can be connected to the computer ports (serial, parallel, USB etc.) or cards connected to the slots in the mother board.

2.2 Stimulus Generation

Variable frequency and variable amplitude sinusoidal waveform is generated using LabVIEW. The waveform is then amplitude modulated with 40 Hz sinusoidal wave [5]. This modulated wave is then played via sound output write and sound output start Vis. Also volume control is done in order to decrease or increase the stimulus sound level thereby getting a variable dB level sound which will further enable us to accurately measure the hearing threshold of the subject.

The octave analysis is done on the sound stimulus which consists of 1/3 fractional analysis with exponential weighting. Thus both the octave graph and total band power is displayed. The frequency of stimulus presented is calculated via tone measurement and displayed on a numeric indicator.

2.3 EEG Acquisition

The EEG signal acquisition is based on the 10-20 lead system. We use an ipsilateral montage. For the recording of EEG, here we use a 3-electrode system. In this system the electrode positions used are Fp, Cz and in any one of the ear lobes as active and reference electrodes. All the electrodes used are simple Ag/AgCl surface cup electrodes and measure the scalp potential.

The electrical activity recorded from scalp is usually in range of $\pm 20\mu\text{V}$, thus for amplification before reaching AD input of data acquisition board, a small battery operated EEG amplifier is used having a gain of 10000 and high pass filter 1Hz and low pass filter of 300Hz.

After the EEG signal is acquired, the fast Fourier spectrum (FFT) is taken and the power spectrum of the signal is displayed. The EEG signal is then filtered using butterworth IIR filter having specifications:

Low pass filter, 3rd order and low cutoff frequency at 40Hz.

Again it is filtered out with IIR smoothing filter. As the EEG waveform is complex in nature, band separation has to be done and then only any special increase or decrease in the amplitude and frequency of those bands can be observed. The four frequency bands are alpha (8-14Hz), beta (14-20Hz), theta (4-8Hz) and delta (0.4-4Hz). In order to do these separations of bands into different frequencies butterworth infinite impulse response filters (IIR) are used as band pass filters as they have better pass band ripple attenuation compared to other classes. The specifications for designing the filter are:

Filter of order 3 having pass band ripple as 1, stop band atten-

uation as 60dB and the upper and lower cutoff frequencies set according to the wave's frequency range mentioned above.

Each of these frequency signals are then fed to the tone measurement express VI which extracts the frequency, amplitude and the phase information of the specific wave at that particular time. These readings can be used to evaluate the response of brain to any stimulus.

2.4 Testing And Analysis

Firstly, the frequency of stimulus is set by frequency control and amplitude control is set to zero, then stimulus is given to subject and simultaneously EEG is recorded and displayed. The physician observes the EEG for AEP waveform in order to get ASSR, especially N1-P1 peaks are searched for. The amplitude control is continuously incremented to increase the volume of the stimulus. The moment the AEP waveform is seen the physician should press the record button, and thus that particular frequency and dB level is updated on the table and the audiogram is simultaneously plotted. Then the frequency control is moved to next frequency on which again the same procedure is performed to get hearing threshold. And thus, the hearing thresholds at all desired frequencies can be found. The automatic analysis provision for determining the type of hearing disorder is also incorporated by comparing each hearing threshold value with standard range and thus an indicator of patient's condition such as normal, mild, severe and profound is provided in the front panel.

3 RESULTS AND DISCUSSIONS

The following results have been obtained and are discussed in this section:

1. The fig. 2 shows the data obtained from recording session by presenting stimuli to both the ears and recording the EEG by varying both amplitude and frequency control.
2. Auditory evoked potentials are only seen in delta waves when stimulus is provided and no change is seen in any other brain wave as shown in fig. 3. As the evoked potential due to sound stimulus are generated and their frequency match to that of delta waves and hence the temporal change in the delta wave is seen.
3. The disorder specific audiograms were taken from different subjects such as the cases ranging from normal, mild, mixed and severe hearing loss were analyzed by the system and the results of which closely adhered to the conventional audiometry. These system recorded audiograms along with the frequency and dB level table are shown in fig. 4, 5 and 6.

4 CONCLUSION

The discussed system is windows based system for real time acquisition and analysis of auditory steady state responses which can be used as objective audiometry.

The system is designed to be flexible and user friendly. The system has simple hardware and software requirements. The

system was successfully tested on subjects with different hearing disorder and the results were found to closely adhere to conventional subjective audiometry.

5 REFERENCE

- [1] Anthony T. Herdman, David R. Stapells and Terence W. Picton. "Place specificity of multiple auditory steady-state responses," *J. Acoustic. Soc. Am.* 112 (4), October 2002.
- [2] Anthony T. Herdman and David R. Stapells. "Thresholds determined using the monotic and dichotic multiple auditory steady-state response technique in normal hearing subjects," *Scandinavian Audiology*; Vol 30; No 1; 2001.
- [3] Hekimoglu, Y., Ozdamar, O. and Delgado, R.E. "Chirp and click evoked auditory steady state response," *IEEETrans.Engineering in medicine and biology*, vol2, no. 2, pp. 2084-2087, November 2002.
- [4] Douglas S. Goodin, Kenneth C. Squires, BeverleyH.Henderson and Arnold Starr. "Age related variations in evoked potentials to auditory stimuli in normal human subjects," *Electroencephalography and Clinical Neurophysiolgy (1978)* Volume 44, Issue: 4, Publisher: Elsevier, Pages: 447-458.
- [5] Anna Van Maanen and David R. Stapells. "Comparison of multiple auditory steady-state responses (80 versus 40 Hz) and slow cortical potentials for threshold estimation in hearing-impaired adult," *International Journal of Audiology* 2005; 44:613/624.
- [6] Small, Susan A. and Stapells, David R. " Multiple auditory steady state response to bone-conduction stimuli to adults with normal hearing," *journal of American academy of audiology*, Volume 16, Number 3, March 2005 , pp. 172-183(12).
- [7] Camila Maia Rabelo and ElianeSchochat. "Sensitivity and specificity o auditory steady-state response testing," *Clinics* 2011; 66(1):87-93.
- [8] Galambos R, Makeig S and Talmachoff PJ. "A 40 Hz auditory potential recorded from the human scalp," *ProcNatlAcadSci USA.* 1981 78: 2643-7, doi: 10.1073/pnas.78.4.2643.
- [9] Picton TW, John MS, Dimitrijevic A and Purcell DW."Human auditory steady-state responses," *Int J Audiol.* 2003; 42:177-219, doi: 10.3109/14992020309101316.
- [10] Van der Werff KR and Brown CJ. "Effect of audiometric configuration on threshold and suprathreshold auditory steady-state responses," *Ear Hear.*2005; 26:310-26, doi: 10.1097/00003446-200506000-00007.
- [11] Aoyagi M, Kiren T, Kim Y, Suzuki Y, Fuse T and Koike Y."Frequency-specificity of amplitude-modulation following response detected by phase spectral analysis," *Audiology* 1993; 32:293±301.
- [12] Aoyagi M, Kiren T, Furuse H, Suzuki Y, Yokota M and Koike Y. "Pure-tone threshold prediction by 80-Hz amplitude-modulation following response," *ActaOtolaryngol (Stockh)* 1994;511:7±14.
- [13] O.G. Lins, T.W. Picton, B. Boucher, A. Durieux-Smith,S.C. Champagne, L.M. Moran, M.C. Perez-Abalo and V.Martin, G. Savio. "Frequency-specific audiometry using steady-state responses," *Ear Hear.* 17 (1996) 81–96.
- [14] G. Rance, R.C. Dowell, F.W. Rickards, D.E. Beer and G.M. Clark." Steady-state evoked potential and behavioral hearing threshold in a group of children with absent click- ABR," *Ear Hear.* 19 (1998) 48–61.
- [15] T.W. Picton, A. Durieux-Smith, S.C. Champagne, J.Whittingham, L.M. Moran, C. Gigue`re and Y. Beauregard." Objective evaluation of aided thresholds using auditory steady-state responses," *J. Am. Acad. Audiol.* 9 (1998)315–331.

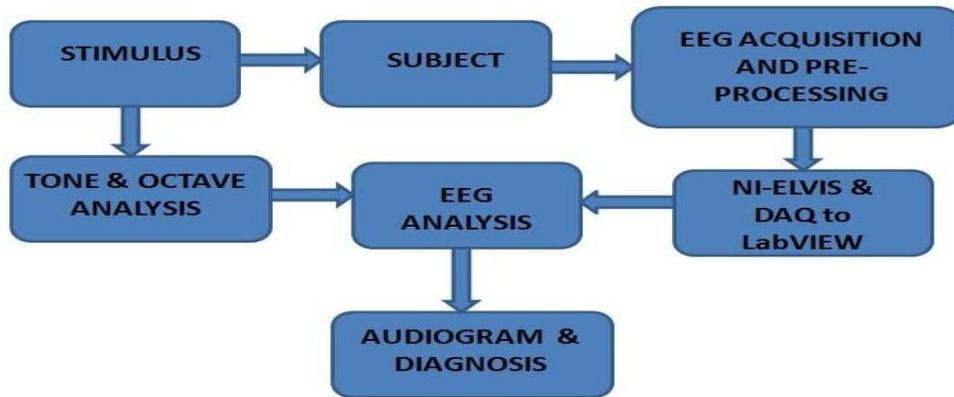


Fig. 1: System representation

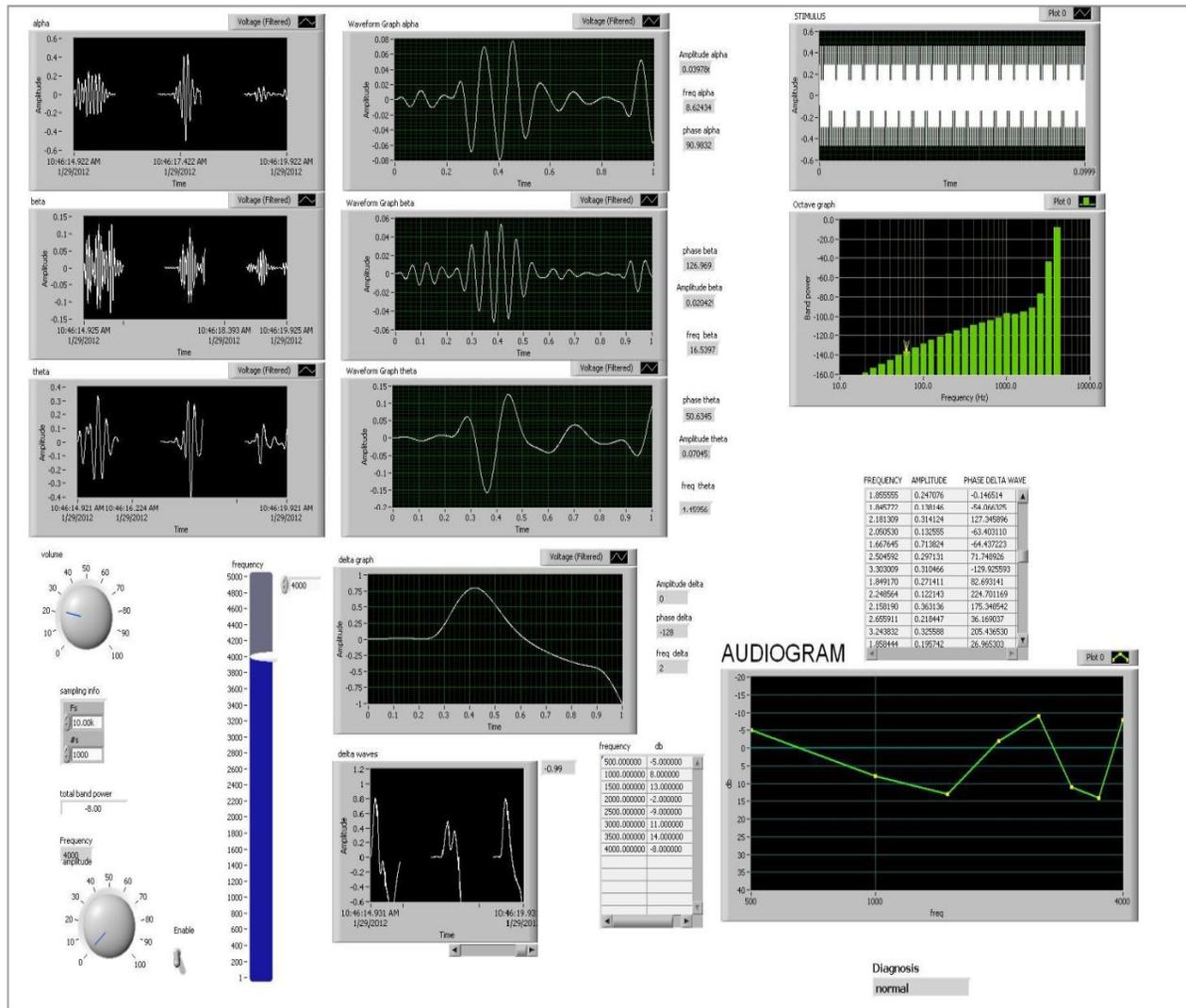


Fig. 2: Audiogram obtained from recording session by providing stimulus on both the ears

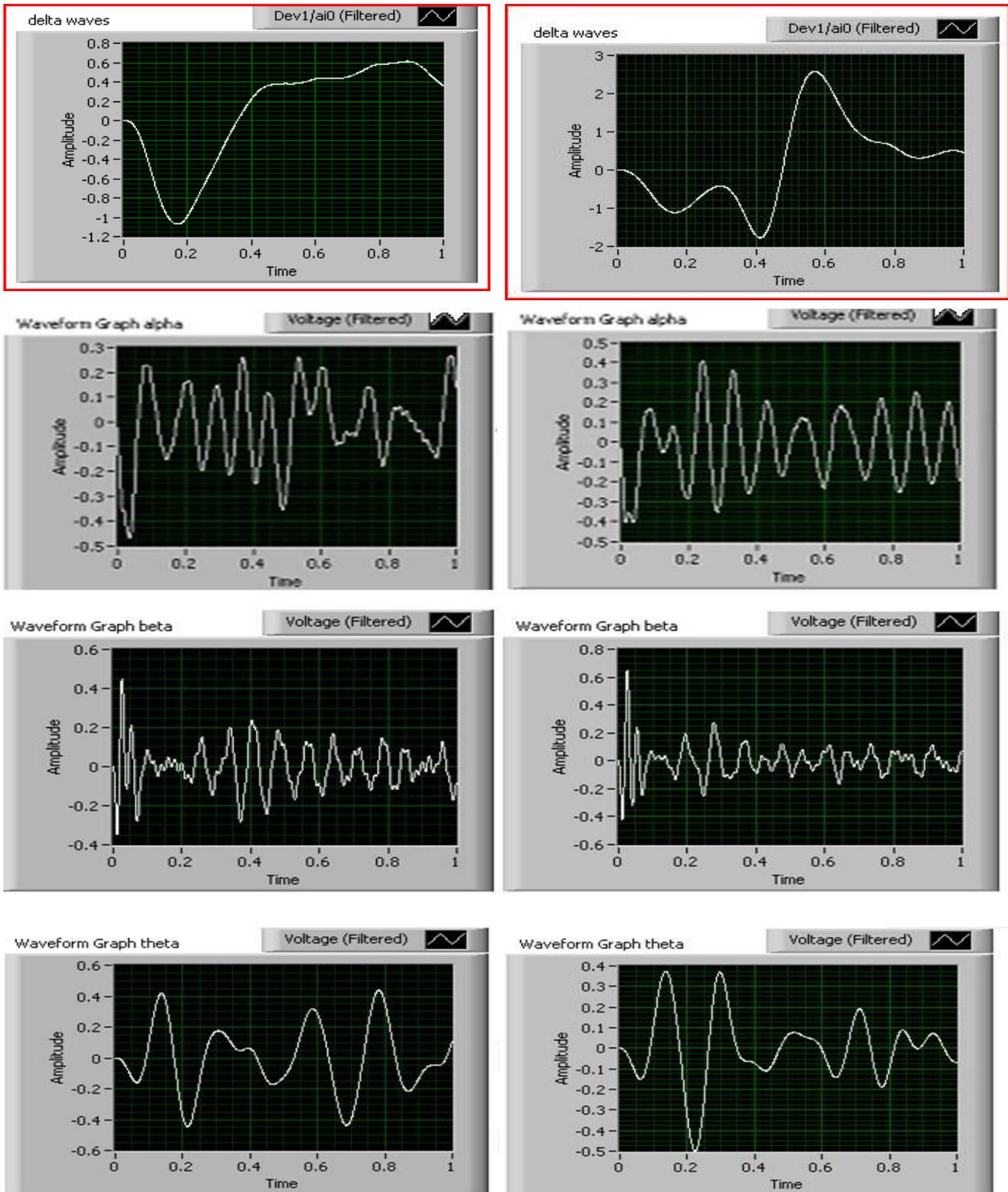


Fig. 3: The brain waves recorded before and after application of stimulus.

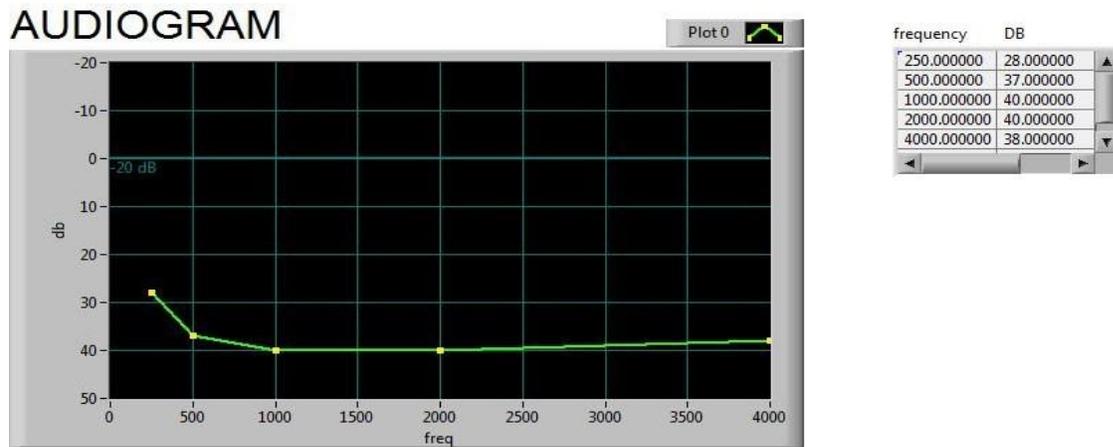


Fig. 4: audiogram obtained for mild loss case

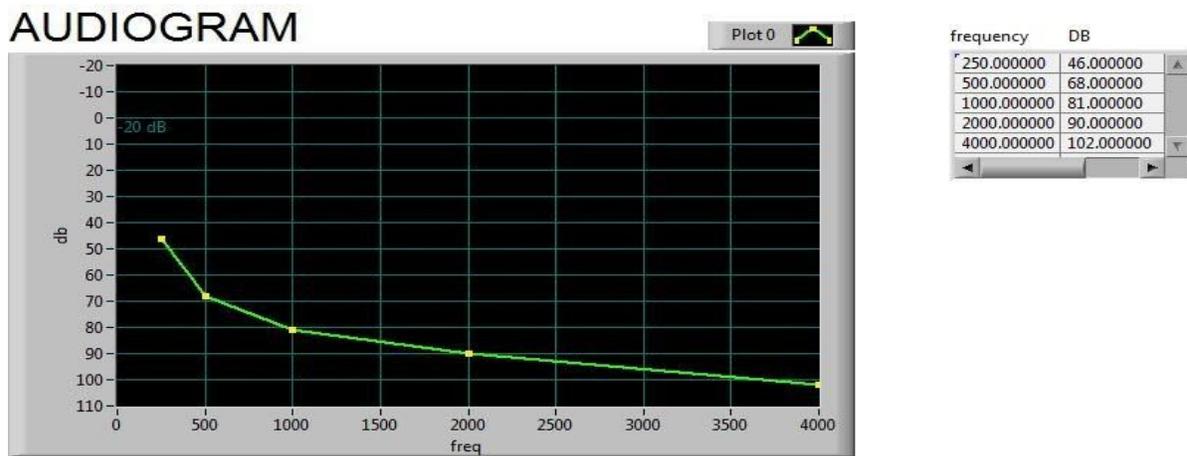


Fig. 5: Audiogram obtained from sensorineural severe loss case

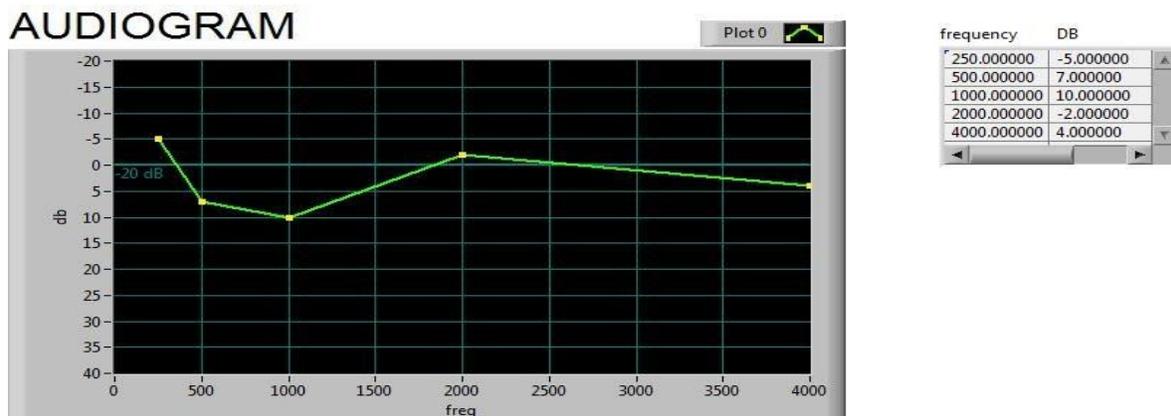


Fig. 6: Audiogram obtained from a normal case.

Evaluation of Feature Extraction and Classification Techniques on Special Symbols

Sanjay S. Gharde, Vidya A. Nemade, K. P. Adhiya

Abstract— The Symbol recognition is a significant area in computer vision that involves the recognition of symbols in an image or video. Symbol recognition is used in a large number of different applications like converting the text document into other electronic document formats, forming mathematical expression, maps, musical notations etc. When someone copies the data from any text document like any PDF file instead of that original characters or special symbols, garbage value would be copied at some places. So identification of special symbols requires higher accuracy at the time of recognition. This paper describes the analytical study of various techniques which will be useful for implementation of the system. Some existing techniques of feature extraction, classification and recognition of symbols and characters are compared in this paper.

Index Terms— Classification, feature extraction, preprocessing, segmentation, symbol recognition, support vector machine,

1 INTRODUCTION

Symbol recognition is used in a large number of diverse applications such as:

- Interpreting and converting scanned engineering drawings and circuit diagrams into other electronic manuscript formats.
- Recognizing and locating trademarked content.
- Querying images from databases based on shape and
- Recognizing characters and words within an electronic document [1].

The existing symbol recognition approaches can be divided usually into two groups: statistical approaches and structural approaches. Normally structural approaches are vector-based in which symbol is decomposed into some vector-based primitives at first such as dominant points, lines and arcs. As for statistical approaches, the pixel is typically the primitive. The most important statistical descriptors include plain binary image, moment invariants, ring projection and shape context [2]. In many applications it is necessary to copy the contents from some original documents which may be in PDF or any other format. So, while copying the data from document, when it encounters a special symbol it remains unread in the copied document, or instead of that special symbol random value placed in the copied document. So it is very difficult to read the original document. While reading the printed English text if a special symbol found then main task is to identify that particular special symbol.

- Sanjay S. Gharde is currently working as assistant professor in Computer department of SSBT College of Engineering and Technology, Jalgaon. PH:+91-9422344964. INDIA, E-mail: sanjay_gharde358@yahoo.com
- Vidya a. Nemade is currently pursuing masters degree in computer science engineering in from North Maharashtra University, Jalgaon, INDIA, PH:+91-9561473161. E-mail: vidya.nemade@gmail.com
- K. P. Adhiya is currently Associate Professor in Computer Engineering Department of SSBT College of Engineering and Technology, Jalgaon, INDIA. PH:+91-9421516499. E-mail: kpadhiya@yahoo.com

After recognition of that symbol same process continues i.e. if again symbol found while reading the rest of the document processing will done on it. Processing of symbol consists of following steps -Preprocessing, Segmentation, Feature extraction, Classification and Recognition.

The architecture of symbol recognition consists of general image processing steps as shown in Fig. 1. Usually image consisting of symbol or character is acquired through scanner or any other digital device. Preprocessing is the next step to produce the clean image after image acquisition and to make it suitable for segmentation and feature extraction. Finally support vector machine or any other suitable classifier can be used for recognition purpose.

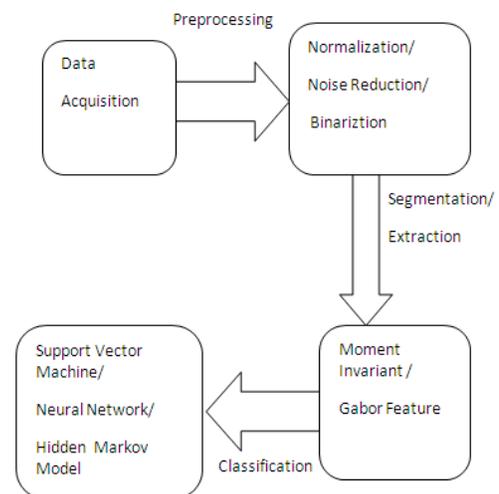


Fig. 1. Architecture of Symbol Recognition

2 RELATED WORK

The history of character recognition can be traced as early as 1900, when the Russian scientist Tyuring attempted to develop an assist for the visually handicapped. The first character recognizers developed in the middle of the 1940s with the

growth of digital computers. The early work on the automatic recognition of characters has been concentrated either upon machine-printed text or upon a small set of handwritten text or symbols. Machine-printed Character recognition systems normally used template matching. For handwritten text, low-level image processing techniques have been used on the binary image to extract feature vectors, which are then fed to statistical classifiers.

Successful, but inhibited algorithms have been implemented mostly for Latin characters and numerals. However, some studies on Japanese, Chinese, Hebrew, Indian, Cyrillic, Greek, and Arabic characters and numerals in both machine-printed and handwritten cases were also initiated. The commercial character recognizers were available in the 1950s. Historical review of character recognition research and growth during this period can be found in and for off-line and on-line cases, respectively. In the early 1990s, image processing and pattern recognition techniques were efficiently united with artificial intelligence (AI) methodologies. Nowadays, in addition to the more powerful computers and more accurate electronic equipments such as scanners, cameras, and electronic tablets, we have efficient, modern use of methodologies such as neural networks (NNs), hidden Markov models (HMMs), fuzzy set reasoning, and natural language processing. [3]

3 PROESS OF SYMBOL RECOGNITION

The symbol recognition normally includes the following parts: Preprocessing, Feature extraction, and Classification.

3.1 Preprocessing

Preprocessing of a character or symbol is important because it affect the recognition rate very much. Preprocessing brings the raw data into some specific format and it gives good results. The main objectives of preprocessing are [3]:

- 1) Noise reduction
- 2) Normalization
- 3) Compression. Some of the existing preprocessing techniques are given in TABLE 1. Recognition system consists of following modules:

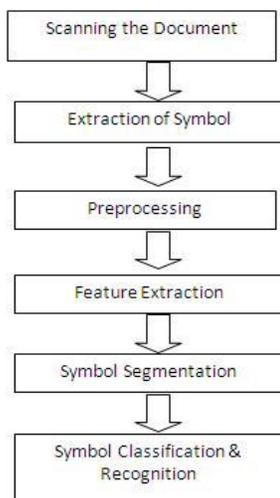


Fig. 2. Process of Symbol Recognition

There are several preprocessing techniques available to produce the clean image. It consists of many steps like Filtering, Morphological operation, Normalization, Thinning, Resampling.

3.1.1 Noise Reduction-

The noise, introduced by the optical scanning device or the writing instrument, causes disconnected line segments and gaps in lines, filled loops, etc. Hundreds of available noise reduction techniques can be categorized in major groups. Some of the existing preprocessing techniques are as follows:

TABLE 1

Existing preprocessing Techniques	
Author	Preprocessing Techniques
Vnaita Mane, Leena Ragma[4]	Filtering, Morphological Operations, Normalization
Birendra Keshari, Stephen Watt[5]	Gaussian Smoothing, Resampling
Gong Xin, LI Cuiyun, PEI Jihong, XIE Weixin[6]	Noise Removal, Smoothing, Data point Rearrangement
Dewi Nasien, Habibollah Haron, Siti Yuharniz[7]	Thinning
Shubhangi D. C, Prof. P.S.Hiremath[8]	Normalization, Thinning
Lie Hu, Richard Zanibbi[9]	Duplicate Point Filtering, Size Normalization, Smoothing, Resampling
J.Pradeep, E. Srinivasan, S.Himavathi, [10]	Binarization, dilation
T. Cheng, J. Khan, H. Liu and D. Y. Y. Yun[11]	autocropping, centralization, and normalization
Birendra Keshari, Stephen Watt[12]	Smoothing, Filling intermediate points, resampling, Size normalization
Widad Jakjoud, Azzeddine Lazrek[14]	Filtering (Gaussian filter) Binarization

3.1.1.1 Filtering:

This aims to remove noise and diminish false points, usually introduced by irregular writing surface and/or poor sampling rate of the data acquisition device. Filters can be planned for smoothing, sharpening, thresholding, removing slightly textured or colored background, and contrast adjustment purposes. [3]

3.1.1.2 Morphological Operations:

The basic idea behind the morphological operations is to filter the document image replacing the convolution operation by

the logical operations [3]. Various morphological operations such as bridge, fill, clean, majority and thin can be intended to connect the broken strokes, decompose the connected strokes, smooth the contours, thin the characters, and extract the boundaries. Therefore, morphological operations can be successfully used to remove the noise on the document images due to low quality of paper and ink, as well as irregular hand movement [4].

3.1.2 Normalization:

Normalization methods try to remove the variations of the writing and acquire standardized data. The following are the basic methods for normalization.

3.1.2.1 Skew Normalization

Due to inaccuracies in the scanning process and writing style, the writing may be slightly tilted or curved within the image. This can harm the effectiveness of later processing and, therefore, should be detected and corrected (e.g., "9" and "g") [3].

3.1.2.2 Size Normalization:

This is used to adjust the character size to a certain standard. Methods of character recognition may apply both horizontal and vertical size normalizations [3].

3.1.2.3 Slant Normalization:

Slant normalization is used to normalize all characters to a standard form. The most common method for slant estimation is the calculation of the average angle of near-vertical elements [3].

3.1.3 Compression:

Compression is mainly used to reduce the storage requirement for image. Two popular compression techniques are thresholding and thinning.

3.1.3.1 Thresholding (Binarization)

In order to reduce storage requirements and to increase processing speed, it is often desirable to represent gray-scale or color images as binary images by picking a threshold value. Two categories of thresholding exist: global and local. Global thresholding picks one threshold value for the entire document image and local thresholding uses the different values for different regions [3].

3.1.3.2 Thinning (Skeletonisation)

This process extracts the shape information of characters or symbols. Skeletonisation is also called thinning. Skeletonisation refers to the process of reducing the width of a line like object from many pixels wide to just single pixel. This process can remove irregularities in letters and in turn, makes the recognition algorithm simpler because they only have to operate on a character stroke, which is only one pixel wide. It also reduces the memory space required for storing the information about the input characters and this process also reduces the processing time [3].

From Table 1 it is observed that Filtering, Normalization, Binarization can be used as a preprocessing techniques.

3.2 Evaluation of Feature Extraction Techniques

Feature extraction is the name given to a family of procedures

TABLE 2
Review of Existing Feature Extraction Methods

Author	No. of Samples	Feature Extraction Techniques
Birendra Keshari, Stephen Watt[5]	Dataset with 48 symbols	Co-ordinates of resampled points, sine & cosines of the angle made by segments and turning angle
Gong Xin, LI Cuiyun, PEI Ji-hong, XIE Weixin[6]	Graphic library contain 110 symbols	Normalizes distance measure, rotation angle of the strokes
Dewi Nasien, Habibollah Haron, Siti Yuhanziz[7]	Lowercase-189,411 samples, Upper-case-217,812 samples	Freeman chain code (4-neighbourhood)
Shubhangi D. C, P.S. Hiremath[8]	26000 Samples of handwritten english character, 5000 samples english handwritten digits	Height, Width, Slantness, Average pseudo pressure, number of Strokes etc.
Lie Hu, Richard Zanibbi [9]	20281 samples in training set, 2202 samples in testing set	Pen up/down, speed, normalized x & y co-ordinate, cosine & sine of curvature, vicinity aspect etc.
J.Pradeep, E. Srinivasan, S.Himavathi [10]	50 dataset	Without any feature Extraction
T. Cheng, J. Khan, H. Liu and D. Y. Y. Yun[11]	34 symbols	Movement invariant
Birendra Keshari, Stephen Watt[12]	137 unique Mathematical symbols	Intersections, loops, Height-width ratio, Number of strokes, Initial / End position etc.
Xue Dong Tian, Hai-Yan Li, Xinfu Li and Li-Ping Zhang[13]	100 Chinese mathematical literature, 3000 mathematical formula	Gabor filter, Parameter selection (standard deviation, wavelength), Elastic meshing
Widad Jakjoud, Azzeddine Lazrek[14]	Symbol & error database	Countour & region approaches
Shaileendra Shrivastava, Sanjay S. Gharde[15]	2000 samples	Moment Invariant and Affine Moment Invariant

for measuring the relevant shape information contained in a pattern so that the task of classifying the pattern becomes easy. The feature extraction stage analyses a text segment and selects a set of features that can be used to uniquely identify the text segment.

The task of human expert is to select features that permit effective and efficient recognition of pattern. Feature extraction is a very significant in recognition system because it is used by the classifier to classify the data. This can be achieved by some of the following:

3.2.1 Chain Code

Chain code is demonstration technique which is useful for image processing, shape analysis and pattern recognition fields. Chain code representation provides the boundary of character image in which the codes represent the direction of where is the location of the next pixel. One such method is Freeman Chain Code having two directions of chain code, namely 4- neighborhood and 8-neighborhood [7].

3.2.2 Moment Invariant: The moment invariants (MIs) are used to evaluate seven distributed parameters of a numeral image [15].

3.2.3 Gabor Filter: 2-D Gabor filter is a somewhat complex sinusoidally modulated Gaussian function with the response in spatial domain and spatial frequency domain.

From Table 2 it is observed that moment Invariant can also be good feature extraction technique because it can be applied to different size images and this approach is invariant to rotation, scaling, and translation and can also be used as aircraft identification, and character recognition. Gabor filter can also be used as feature extraction because along with this Xue-Dong Tian [13] achieved high recognition rate.

3.3 Evaluation of Classification Techniques

During classification stage character or symbol is placed in the appropriate class to which it belongs. Various classifiers can be used for the recognition purpose. Some of them are: Support Vector Machine (SVM), Neural Network (NN), Hidden Markov Model (HMM)

3.3.1 Support Vector Machine

Generally, Support Vector Machines (SVM) is used for classification in pattern recognition. Support Vector machine is one of the supervised learning technique. First realistic implementation of SVM had been executed in early nineties. This strategy is introduced by Vapnik and co-workers. Support vector machine is one of the paramount techniques used for linear and nonlinear classification. The SVM classifier was formerly developed for two-class or binary classification and the challenging applications of pattern recognition led to the design of multi-class SVM classifiers using the binary SVM classifiers.

SVMs were developed to solve the classification problem, but recently they have been unlimited to solve regression problems [15].

3.3.2 Hidden Markov Model

The research work of Hidden Markov Model can be traced back to 1960's, but the HMM is not generally used in the pattern recognition field until 1980's. This model is primarily applied in the continuous voice recognition and great success has been gained. In recent years, HMM is also be used in the handwriting recognition and cursive script recognition.

Hidden markov model is a dual random process each comes from a markov process. One of these random process is not obvious, its characteristic can only he described by the other random process's surveillance. This hidden process is a finite state process [6].

3.3.3 Neural Network

Feedforward neural networks, including multilayer perceptron (MLP), radial basis function (RBF) network, the probabilistic neural network (PNN), higher-order neural complex (HONN), etc., have been broadly applied to pattern recognition. The connecting weights are usually adjusted to minimize the squared error on training samples in supervised learning. A network using local connection and shared weights, called convolutional neural network, has reported huge success in character recognition. Thers is a Zernike moment feature based approach for Devnagari handwritten character recognition. They used an artificial neural network for classification.

TABLE 3
Evaluation of Classification method with their Recognition Rate

Author	Classifier	Rec. Rate
Gong Xin, LI Cuiyun, PEI Jihong, XIE Weixin[6]	Hidden Markov Model	85.0%
Shubhangi D. C, Prof. P.S.Hiremath[8]	Multi class SVM Classifier	99.91%
Lie Hu, Richard Zanibbi[9]	Hidden Markov Model	92.31%
J.Pradeep, E. Srinivasan, S.Himavathi, [10]	Feed forward backpropogation Neural Network	90.19%
Xue-Dong Tian, Hai-Yan Li, Xin-Fu Li and Li-Ping Zhang[13]	Minimum distance Classifier	97.75%
Shailedra Shrivastava, Sanjay S. Gharde[15]	Support Vector Machine	99.48%
B.Q.Huang, M.T. Kechadi[16]	Hidden Markov Model & Multi-layer Perceptron	93.43% (Avg.)
C. Pirlo, D. Impedovo[17]	Zoing Based Classification	95%
M. Hanmandlua, K.R. Murali Mohanb, Sourav Chakrabortyc, Sumeer Goyald, D. Roy Choudhurye	Neural Network, Fuzzy Logic	93.84% (NN) 98.25% (Fuzzy Logic)

From Table 3 it is observed that Gong [6] obtained a recognition rate of 85% using hidden markov model and Haung [16] obtained the recognition rate of 93.43%. Support vector machine [8], [15] provides the recognition rate upto 99.91% to 99.48%. Using minimum distance classifier Xue-Dong Tian [13] obtained the recognition rate 95.75%. C. Pirlo [17] proved that Zoing Based Classification provides the recognition rate of 95%. So it is recognized that recognition rate can be higher using support vector machine.

4 CONCLUSION

There are various approaches for recognition of symbols as neural network, support vector machine and hidden markov model etc. After review study it is observed that support vector machine can provides better recognition rate as compared to other classification techniques.

From the evaluation of feature extraction and classification techniques, system can be implemented for recognition of special symbol using moment invariant as a feature extraction technique and support vector machine as a classifier.

REFERENCES

- [1] Alexander Wong and William Bishop, Robust Hough-Based Symbol Recognition Using Knowledge-Based Hierarchical Neural Networks.
- [2] Su Yang, "Symbol Recognition via Statistical Integration of Pixel-Level Constraint Histograms: a New Descriptor" *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol 27, No. 2, 278-281, Feb 2005.
- [3] Nafiz Arica and Fatos T. Yarman-Vural, "An Overview of Character Recognition Focused on of Character Recognition Focused on Off-Line Handwriting" 2001 IEEE.
- [4] Vanita Mane, Leena Ragha, "Handwritten Character Recognition using Elastic Matching and PCA" International Conference on Advances in Computing, Communication and Control (ICAC3'09) 2009 ACM, 410-415, 978-1-60558-351-8.
- [5] Birendra Keshari and Stephen Watt, "Online Mathematical Symbol Recognition using SVM's with Features from Functional Approximation
- [6] Gong Xin, LI Cuiyun, PEI Jihong, XIE Weixin, "HMM Based Online Hand-Drawn Graphics Symbol Recognition" ICSP'02 Proceedings 1067-1070, 0-7803-7488-6, 2002 IEEE.
- [7] Dewi Nasien, Habibollah Haron, Siti Yuhaniz, "Support Vector Machine (Svm) For English Handwritten Character Recognition", 2010 Second International Conference on Computer Engineering and Applications, 249-252, 978-0-7695-3982-9/10 2010 IEEE.
- [8] Shubhangi D. C, Prof. P.S.Hiremath, "Handwritten English Character And Digit Recognition Using Multiclass SVM Classifier And Using Structural Micro Features", International Journal of Recent Trends in Engineering, Vol 2, No. 2, November 2009, 193-195.
- [9] Lie Hu, Richard Zanibbi, "HMM-Based Recognition of Online Handwritten Mathematical symbols Using Segmental K-means Initialization and A Modified Pen-up/down Feature" 2011 International conference on Document analysis and Recognition, 457- 462, 1520-5363/11, 2011 IEEE.
- [10] J.Pradeep, E. Srinivasan, S.Himavathi, "Neural Network based Handwritten Character Recognition system without feature extrac-

- tion" International Conference on Computer, Communication and Electrical Technology ICCET March, 2011, 40-44, 2011 IEEE
- [11] T. Cheng, J. Khan, H. Liu and D. Y. Y. Yun, "A Symbol recognition system", 0-81864960-7193, 918-920, 1993 IEEE.
- [12] Birendra Keshari and Stephen M. Watt, "Hybrid Mathematical Symbol Recognition using Support Vector Machines".
- [13] Xue-Dong Tian, Hai-Yan Li, Xin-Fu Li and Li-Ping Zhang, "An Improved Method Based On Gabor Feature For Mathematical Symbol Recognition", *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August 2007*, 1678-1682, -4244-0973-X/07 2007 IEEE.
- [14] Widad Jakjoud, Azzeddine Lazrek, "Segmentation method of offline Mathematical symbol", 978-1-61284-732-0/11 2010 IEEE.
- [15] Shaileendra Kumar Shrivastava, Sanjay S. Gharde, "Support Vector Machine for Handwritten Devanagari Numeral Recognition" International Journal of Computer Applications (0975 – 8887), 9-14, Volume 7– No.11, October 2011.
- [16] B.Q.Huang, M.T. Kechadi, "A Fast Feature Selection Model for Online Handwriting Symbol Recognition".
- [17] G. Pirlo and D. Impedovo, "Fuzzy-Zoning-Based Classification for Handwritten Characters", *IEEE Transactions On Fuzzy Systems*, 780-785, VOL. 19, NO. 4, AUGUST 2011, 1063-6706.
- [18] M. Hanmandlua, K.R. Murali Mohanb, Sourav Chakrabortyc, Sumeer Goyald, D. Roy Choudhurye, "Unconstrained handwritten character recognition based on fuzzy logic" 0031-3203/02, 603-623, Pattern Recognition 2003.

Frequent Item set Generation by Parallel Preprocessing on Generalized Dataset

Gurudatta Verma, Vinti Nanda

ABSTRACT- One of the important problems in data mining is discovering association rules from databases of transactions where each transaction consists of a set of items. Many industries are interested in developing the association rules from their databases due to continuous retrieval and storage of huge amount of data. The discovery of interesting association relationship among business transaction records in many business decision making process such as catalog decision, cross-marketing, and loss-leader analysis. The most time consuming operation in this discovery process is the computation of the frequency of interesting subset of items (called candidates) in the database of transactions. Hence, it is has become vital to develop a method that may avoid or reduce candidate generation and test and utilize some novel data structures to reduce the cost in frequent pattern mining. In this paper, We have proposed An Integrated approach of Parallel Computing and ARM for mining Association Rules in Generalized data set that is fundamentally different from all the previous algorithms in that it uses database in transposed form and database transposition is done using Parallel transposition algorithm (Mesh Transpose) so to generate all significant association rules number of passes required is reduced. We will compare proposed algorithm with Apriori algorithm for frequent item sets generation. The CPU and I/O overhead can be reduced in our proposed algorithm and it is much faster than other Association Rule Mining algorithms.

Index Words-Data Mining, Association Rule Mining (ARM), Association rules, Apriori algorithm, Frequent pattern.



1 INTRODUCTION

The rapid development of computer technology, especially increased capacities and decreased costs of storage media, has led businesses to store huge amounts of external and internal information in large databases at low cost. Mining useful information and helpful knowledge from these large databases has thus evolved into an important research area [3, 2, 1].

Association rule mining (ARM) [18] has become one of the core data mining tasks and has attracted tremendous interest among data mining researchers. ARM is an undirected or unsupervised data mining technique which works on variable length data, and produces clear and understandable results. Association Rule Mining (ARM) algorithms [17] are defined into two categories; namely, algorithms respectively with candidate generation and algorithms without candidate generation. In the first category, those algorithms which are similar to Apriori algorithm for candidate generation are considered. Eclat may also be considered in the first category [8]. In the second category, the FP-Growth algorithm is the best-known algorithm.

The main drawback of earlier algorithms is the repeated scans over large database. This may be a cause of decrement in CPU performance, memory and increment in I/O overheads. The performance and efficiency of ARM algorithms mainly depend on three factors; namely candidate sets generated, data structure used and details of implementations [8]. In this paper we have proposed an Algorithm which uses these three factors. Suppose if there are 104 frequent 1 itemsets, Apriori algorithm may produce 107 candidate 2 itemsets, count them and judge their frequency [11]. Besides, it may produce as many as 2100 (about 1030) candidate itemsets in order to find the frequent itemset which includes 100 items. What's more, it may scan the database many times to check a larger candidate through matching mode. Table 1 represents the comparison between the apriori and our algorithm.

Table1: Comparison of Apriori with Our Algorithm

Algorithm	Data Preprocessing	Scan	Data Set
Apriori	No Facility	Repeated Scan	Boolean
Our Algorithm	Parallel Preprocessing	One Time Scan	Generalized

- Gurudatta Verma is currently pursuing Masters degree program in computer science and engineering from Chhatrapati Shivaji Institute of Technology, Durg(C.G),India,E-mail:gurudatta.verma@gmail.com
- Vinti Nanda is an assistant professor in the Department of computer science and engineering,Chaatrapati Shivaji Institute of Technology, Durg(C.G.), India. E-mail:vintinanda@csitdurg.in

Transactional database is considered as a two dimension array which works on generalized value dataset. The main difference between proposed algorithm and other algorithms is that instead of using transactional array in its natural form,

our algorithm uses transpose of array i.e. rows and columns of array are interchanged and transposition is done using parallel matrix transpose algorithm (Mesh Transpose) [20]. The parallel architecture that lends itself most naturally to matrix operations is the mesh. Indeed, an $n \times n$ mesh of processors can be regarded as a matrix and is therefore perfectly fitted to accommodate an $n \times n$ data matrix, one element per processor. This is precisely the approach we shall use to compute the transpose of an $n \times n$ matrix initially stored in an $n \times n$ mesh of processors. We find that the time taken for matrix transpose decreases with an increase in the number of processors. We also observe that the speedup is very high for small as well as very large size of matrix when we increase the number of processors. The idea of our algorithm is quite simple. Since the diagonal elements are not affected during the transposition, that is, element a_{ii} of A equals element a_{ii} of AT , the data in the diagonal processors will stay stationary.

The advantage of using transposed array is to calculate support count for particular item. There is no need to repeatedly scan array. Only by finding the row sum of the array will give the required support count for particular item, which ultimately results in increased efficiency of the algorithm. The major Advantages of our proposed algorithm are as follows:-

- Candidate generation becomes easy and fast.
- Association rules are produced much faster, since retrieving a support of an itemset is quicker.
- The original file isn't influenced by the pruning process where its role ends as soon as data is stored in 2-d array.
- It reduces the I/O overhead
- The retrieval of support of an itemset is quicker

The remainder of this paper is organized as follows: Section 2 provides a brief review of the related work. In Section 3, we explain Frequent Itemset and Association Rule Mining through Apriori Algorithm. In Section 4, we introduce our approach of frequent itemset generation using parallel preprocessing. An illustration of the algorithm and experiment analysis is presented in section 5 and section 6 respectively. Finally, we concluded our work.

2 RELATED WORK

One of the most well known and popular data mining techniques is the Association rules or frequent item sets mining algorithm. The algorithm was originally proposed by Agrawal et al. [4] [5] for market basket analysis. Because of its significant applicability, many revised algorithms have been introduced since then, and Association rule mining is still a widely researched area.

Agrawal et. al. presented an AIS algorithm in [4] which generates candidate item sets on-the-fly during each pass of the database scan. Large item sets from previous pass are

checked if they are present in the current transaction. Thus new item sets are formed by extending existing item sets. This algorithm turns out to be ineffective because it generates too many candidate item sets. It requires more space and at the same time this algorithm requires too many passes over the whole database and also it generates rules with one consequent item.

Agrawal et. al. [5] developed various versions of Apriori algorithm such as Apriori, AprioriTid, and AprioriHybrid. Apriori and AprioriTid generate item sets using the large item sets found in the previous pass, without considering the transactions. AprioriTid improves Apriori by using the database at the first pass. Counting in subsequent passes is done using encodings created in the first pass, which is much smaller than the database. This leads to a dramatic performance improvement of three times faster than AIS.

Scalability is another important area of data mining because of its huge size. Hence, algorithms must be able to "scale up" to handle large amount of data. Eui-Hong et. al [16] tried to make data distribution and candidate distribution scalable by Intelligent Data Distribution (IDD) algorithm and Hybrid Distribution (HD) algorithm respectively. IDD addresses the issues of communication overhead and redundant computation by using aggregate memory to partition candidates and move data efficiently. HD improves over IDD by dynamically partitioning the candidate set to maintain good load balance. Different works are reported in the literature to modify the Apriori logic so as to improve the efficiency of generating rules. These methods even though focused on reducing time and space, in real time still needs improvement.

3 FREQUENT ITEM SET AND ASSOCIATION RULE

The aim of Association rule mining is exploring relations and important rules in large datasets. A dataset is considered as a sequence of entries consisting of attribute values also known as items. A set of such item sets is called an item set. Frequent item sets are sets of pages which are visited frequently together in a single server session.

Let $I = \{ I_1, I_2, \dots, I_m \}$ be a set of items. Let D , the task-relevant data, be a set of database transactions where each transaction T is a set of items such that $T \subseteq I$. Each transaction is associated with an identifier, called TID. Let A be a set of items. A transaction T is said to contain A if and only if $A \subseteq T$. An association rule is an implication of the form $A \Rightarrow B$, where $A \subseteq I$, $B \subseteq I$, and $A \cap B = \emptyset$. The rule $A \Rightarrow B$ holds in the transaction set D with support s , where s is the percentage of transactions in D that contain $A \cup B$ (i.e., the union of sets A and B , or say, both A and B). This is taken to be the probability, $P(A \cup B)$. The rule $A \Rightarrow B$ has confidence c in the transaction set D , where c is the percentage of transactions in D containing A that also contain B . This is taken to be the conditional probability, $P(B|A)$. That is,

$$\text{support}(A \Rightarrow B) = P(A \cup B) \dots \dots \dots (2.1)$$

$$\text{confidence}(A \Rightarrow B) = P(B | A) \dots \dots \dots (2.2)$$

A set of items is referred to as an itemset. An itemset that contains k items is a k-itemset. The set {bread, butter} is a 2-itemset. The occurrence frequency of an itemset is the number of transactions that contain the itemset, it is also known, as the frequency, or support count. If the relative support of an itemset I satisfies a pre specified minimum support threshold then I is a frequent itemset. The set of frequent k-itemsets is commonly denoted by L_k. From Equation (2.2), we have

$$\text{confidence}(A \Rightarrow B) = \frac{P(B|A)}{P(A)} = \frac{\text{support}(A \cup B)}{\text{support}(A)} = \frac{\text{support_count}(A \cup B)}{\text{support_count}(A)} \dots \dots \dots (2.3)$$

Let $\tau = I_1, I_2, \dots, I_m$ be a set of binary attributes, called items. Let T be a database of transactions. Each transaction t is represented as a binary vector, with $t[k] = 1$ if t bought the item I_k , and $t[k] = 0$ otherwise. There is one tuple in the database for each transaction. Let X be a set of some items in τ . We say that a transaction t satisfies X if for all items I_k in X, $t[k] = 1$.

By an association rule, we mean an implication of the form $X \Rightarrow I_j$, where X is a set of some items in τ , and I_j is a single item in τ that is not present in X. The rule $X \Rightarrow I_j$ is satisfied in the set of transactions T with the confidence factor $0 \leq c \leq 1$ if at least c% of transactions in T that satisfy X also satisfy I_j . We will use the notation $X \Rightarrow I_j | c$ to specify that the rule $X \Rightarrow I_j$ has a confidence factor of c.[3]

3.1 Apriori Algorithm

The Apriori algorithm is one of the most popular algorithms for mining frequent patterns and association rules [4]. It introduces a method to generate candidate itemsets C_k in the pass k of a transaction database using only frequent itemset L_{k-1} in the previous pass. The idea rests on the fact that any subset of a frequent itemset must be frequent as well. Hence, C_k can be generated by joining two itemsets in L_{k-1} and pruning those that contain any subset that is not frequent as shown in Fig1.

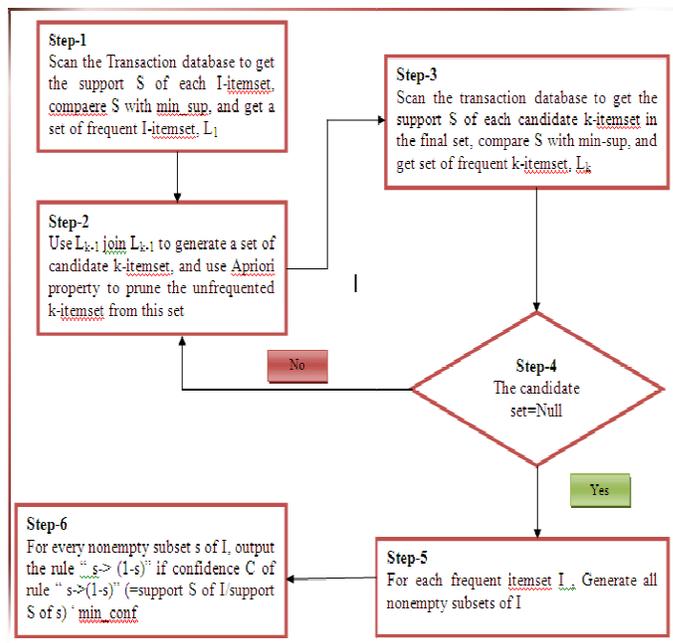


Figure1 : Apriori Algorithm

4 OUR APPROACH

Association rule and frequent itemset mining has become now a widely research area and hence, faster and faster algorithms have been presented. The Association Rule Mining algorithms such as Apriori, FP-Growth requires repeated scans over the entire database. All the input/output overheads that are being generated during repeated scanning the entire database decrease the performance of CPU, memory and I/O overheads.

Much work has been carried out on improving the efficiency of the apriori algorithm by reducing the I/O time and minimizing the set of candidate itemsets. However, all these works suffer from problem of scans over the database at least once. The efficiency of these algorithms can still be improved by reducing the time required for counting the supports of candidate itemsets. We aim to obtain an efficient algorithm which reduces the time needed to count the supports of candidate itemsets.

4.1 Parallel Data Preprocessing

The idea of our algorithm is quite simple. Since the diagonal elements are not affected during the transposition, that is, element a_{ii} of A equals element a_{ii} of AT, the data in the diagonal processors will stay stationary. Those below the diagonal are sent to occupy symmetrical positions above the diagonal (solid arrows in Fig. 2). Simultaneously, the elements above the diagonal are sent to occupy symmetrical positions below the diagonal (dashed arrows in Fig.2).

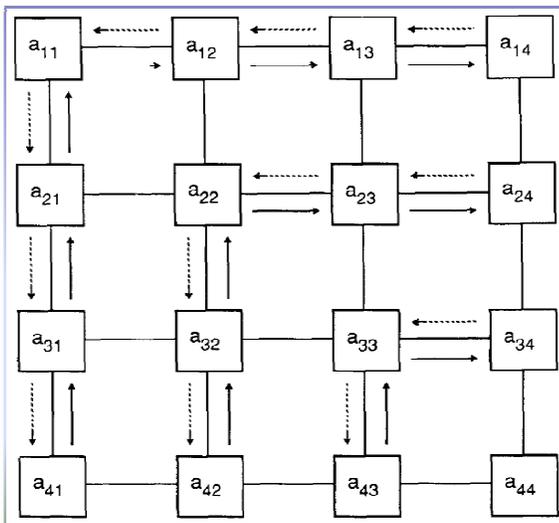


Figure 2. Matrix to be transposed, stored in mesh of processors.

4.2 Our Proposed Algorithm

We propose a new algorithm called in which Transactional database is considered as a two dimension array which works on generalized value dataset. The main difference between proposed algorithm and other algorithms is that instead of using transactional array in its natural form, our algorithm uses transpose of array i.e. rows and columns of array are interchanged and transposition is achieved using parallel matrix transpose algorithm.

```

Procedure EPTA()
// Transpose the transactional database
1. Transpose(Data Set)
2. Read the database to count the support of C1 to determine
L1 using sum of rows.
3. L1= Frequent 1- itemsets and k:= 2
4. While (k-1 ≠ NULL set) do
Begin
Ck := Call Gen_candidate_itemsets (Lk-1)
Call Prune (Ck)
for all itemsets i ∈ I do
Calculate the support values using dot-multiplication of
array;
Lk := All candidates in Ck with a minimum support;
k:=k+1
End
5. End of step-4
End Procedure
    
```

```

Procedure TRANSPOSE (A)
Step 1: do steps 1.1 and 1.2 in parallel
(1.1) for i = 2 to n do in parallel
for j = 1 to i - 1 do in parallel
C(i- 1, j) (a, j, i)
end for
end for
(1.2) for i = 1 to n -1 do in parallel
for j = i + 1 to n do in parallel
B(i, rj -1) (ai, j, i)
end for
end for
Step 2: do steps 2.1, 2.2, and 2.3 in parallel
(2.1) for i = 2 to n do in parallel
for j = 1 to i - 1 do in parallel
while P(i, j) receives input from its
neighbors do
(i) if (aki, m, k) is received from P(i +
L, j)
then send it to P(i - 1, j)
end if
(ii) if (ak, m, k) is received from P(i -
L, j)
then if i = m and j = k
then A(i, j) - a, {ak, has
reached its destination}
else send (aki, m, k) to P(i
+ L, j)
end if
end if
end while
end for
(2.2) for i = 1 to n do in parallel
while P(i, i) receives input from its neighbors do
(i) if (ak, m, k) is received from P(i + L, i)
then send it to P(i, i + 1)
end if
(ii) if (ak, m, k) is received from P(i, i + 1)
then send it to P(i + L, i)
end if
end while
end for
(2.3) for i = 1 to n - 1 do in parallel
for j = i + 1 to n do in parallel
while P(i, j) receives input from its neighbors do
(i) if (ak, m, k) is received from P(i, j + 1)
then send it to P(i, j - 1)
end if
(ii) if (akin, m, k) is received from P(i,
j -1)
then if i = m and j = k
then A(i, j) +- ak. {ak, has
reached its destin
ation}
else send (aki, m, k) to P(i,
j + 1)
end if
end if
end while
end for
end for.
End Procedure//TRANSPOSE
    
```

```

Procedure Gen_candidate_itemsets (Lk-1)
    Ck =  $\Phi$ 
    for all itemsets I1 ∈ Lk-1 do
    for all itemsets I2 ∈ Lk-1 do
    if I1[1] = I2[1] ^ I1[2] = I2[2] ^ ... ^ I1[k-1] < I2[k-1] then
        c = I1[1], I1[2] ... I1[k-1], I2[k-1]
    Ck = Ck ∪ {c}
    End Procedure
    
```

```

Procedure Prune(Ck)
    for all c ∈ Ck
    for all (k-1)-subsets d of c do
    if d ∉ Lk-1
    then Ck = Ck - {c}
    End Procedure
    
```

5 AN ILLUSTRATION

Suppose we have a transactional database in which the user transactions from T1 to T5 and items from A1 to A5 are stored in the form of generalized values, which is shown in Table-1(Fig 3)

Consider the transpose of transactional database as shown in Table-1 is stored in Table-2 by applying Parallel Transposition that can be used in our proposed algorithm. Assume the user specified minimum support is 40%, and then the steps for generating all frequent item sets in proposed algorithm will be repeated until NULL set is reached. In our algorithm, transactional dataset will be used in the transposed form. Therefore, candidate set and frequent itemset generation process will be changed as compared to Apriori algorithm.

Then the candidate 2-itemset will be generated by performing dot-multiplication of rows of array, as array consist of generalized values, the resultant cell will be produce in the form of 1. If the corresponding cells of the respective rows have 1, otherwise 0 will be in the resultant cell. In this approach, we will receive a new array consisting of candidate 2-itemsets to get the higher order of itemsets. The above process between rows of array can be performed to find out the results.

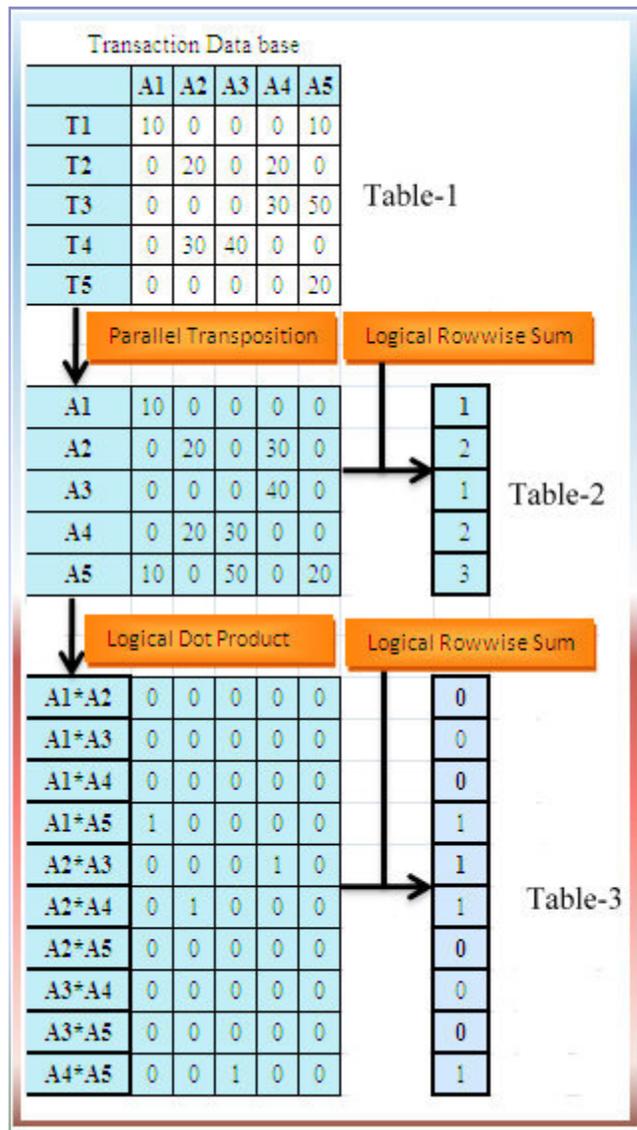


Figure 3. An Illustration of our approach

6 EXPERIMENTAL EVALUATIONS

The performance comparison of our mining algorithm with classical frequent pattern-mining algorithm Apriori is shown in Fig 4 . All the experiments are performed on 1.50Ghz Pentium-iv desktop machine with 256 MB main memory, running on Windows-XP operating system. The program for Apriori and our proposed algorithm were developed in Java JDK1.5 environment.

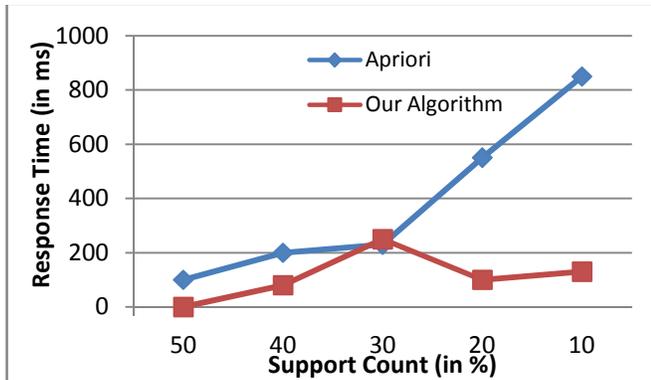


Figure 4. Performance analysis of algorithms.

7 CONCLUSIONS

ARM algorithms are important to discover frequent itemsets and patterns from large databases. In this project, we have designed An Algorithm for generation of frequent itemsets similar to Apriori algorithm. The proposed algorithm can improve the efficiency of Apriori algorithm and it is observed to be very fast. Our algorithm is not only efficient but also very fast for finding association rules in large databases. The proposed algorithm drastically reduces the I/O overhead associated with Apriori algorithm and retrieval of support of an itemset is quicker as compared to Apriori algorithm. This algorithm may be useful for many real-life database mining scenarios where the data is stored in generalized form. Our algorithm uses parallel transposition of generalized 2D data set, so the data preprocessing goes faster. This algorithm cannot be used with multimedia dataset.

REFERENCES

- [1] C.-Y. Wang, T.-P. Hong and S.-S. Tseng. "Maintenance of discovered sequential patterns for record deletion". *Intell. Data Anal.* pp. 399-410, February 2002.
- [2] M.S. Chen, J.Han and P.S. Yu. "Data Mining : An overview from a database perspective", *IEE Transactions on Knowledge and Data Engineering* 1996.
- [3] R.Agrawal, T. Imielinski and A. Swami, "Database Mining: a performance perspective", *IEE Transactions on knowledge and Data Engineering*, 1993.
- [4] Agrawal, R., Imielinski, T., and Swami, A. N. "Mining Association Rules Between Sets of Items in Large Databases". *Proceedings of the ACM SIGMOD, International Conference on Management of Data*, pp.207- 216, 1993.
- [5] Agrawal. R., and Srikant. R., "Fast Algorithms for Mining Association Rules", *Proceedings of 20th International Conference of Very Large Data Bases.* pp.487-499,1994.
- [6] Jong Park, S., Ming-Syan, Chen, and Yu, P. S. "Using a Hash-Based Method with transaction Trimming for Mining Association Rules". *IEEE Transactions on Knowledge and Data Engineering*, 9(5), pp.813-825,1997.
- [7] M.H.Margahny and A.A.Mitwaly, "Fast Algorithm for Mining Association

- Rules" in the conference proceedings of AIML, CICC, pp(36-40) Cairo, Egypt, 19-21 December 2005.
- [8] Y.Fu., "Discovery of multiple-level rules from large databases", 1996.
- [9] F.Bodon, "A Fast Apriori Implementation", in the Proc.1st IEEE ICDM Workshop on Frequent Itemset Mining Implementations (FIMI2003, Melbourne,FL).CEUR Workshop Proceedings 90, A acheme, Germany 2003.
- [10] Akhilesh Tiwari, Rajendra K. Gupta, and Dev Prakash Agrawal, "Cluster Based Partition Approach for Mining Frequent Itemsets" in the *International Journal of Computer Science and Network Security(IJCSNS)*, VOL.9 No.6,pp(191-199) June 2009.
- [11] JiaWei Han Micheline Kamber."Data Mining:Concepts and Techniques"[M].Translated by Ming FAN, XaoFeng MENG etc. mechanical industrial publisher,Beijing,2001,150-158.
- [12] M.J. Zaki. "Scalable algorithms for association mining". *IEEE Transactions on Knowledge and Data Engineering*, 12 : 372 -390, 2000.
- [13] Jochen Hipp, Ulrich Guntzer, Gholamreza Nakhaeizadeh. "Algorithms for Association Rule Mining - A General Survey and Comparison".*ACM SIGKDD*, July 2000, Vol-2, Issue 1, page 58-64.
- [14] Sotiris Kotsiantis, Dimitris Kanellopoulos. "Association Rules Mining: A Recent Overview". *GESTS International Transactions on Computer Science and Engineering*, Vol.32 (1), 2006, pp. 71-82.
- [15] S. Brin, R. Motwani, J. D. Ullman, AND S. Tsur, "Dynamic itemset counting and implication rules for market basket data", *SIGMOD Record* 26(2), pp. 255-276, 1997.Kim Man Lui, Keith C.C. Chan, and John Teofil Nosek "The Effect of Pairs in Program Design Tasks" *IEEE transactions on software engineering*, VOL. 34, NO. 2, march/april 2008.
- [16] Eui-Hong Han, George Karypis, and Kumar, V. Scalable "Parallel Data Mining for Association Rules". *IEEE Transaction on Knowledge and Data Engineering*, 12(3), pp.728-737, 2000.
- [17] Sanjeev Kumar Sharma & Ugrasen Suman "A Performance Based Transposition Algorithm for Frequent Itemsets Generation" *International Journal of Data Engineering (IJDE)*, Volume (2) : Issue (2) : 2011
- [18] Dr (Mrs).Sujni Paul "An Optimized Distributed Association Rule Mining Algorithm In Parallel and Distributed Data Mining With Xml Data For Improved Response Time".*International Journal Of Computer Science And Information Technology*, Volume 2, Number 2, April 2010
- [19] Manoj Bahel, Chhaya Dule "Analysis of frequent item set generation process in Apriori & RCS (Reduced Candidate Set) Algorithm" *National Conference on Information and Communication Technology*, Bangalore April 2010
- [20] Sedukhin, S.G.; Zekri, A.S.; Myiazaki, T."Orbital Algorithms and Unified Array Processor for Computing 2D Separable Transforms" *Parallel Processing Workshops (ICPPW)*, 2010 39th International Conference Page(s): 127 - 134

An SVD Based Approach for the Removal of Artifacts from the Images Compressed by BDCT

Shanty Chacko, Greeshma K.S, J.Jayakumar

Abstract - Blocking artifacts continue to be among the most serious defects that occur in images and video streams compressed to low bit rates using block discrete cosine transform (BDCT) (e.g., JPEG, MPEG and H.263). In the proposed scheme a signal adaptive, space invariant & nonlinear filter is used to smoothen the DC sub image. The compressed image is divided into monotone & edge blocks based on the ac components present in each block. Then a Singular Value Decomposition filter is applied to the non edge blocks to remove the artifacts in the block. In the edge blocks the proposed filter is not applied since the artifacts in the blocks are not sensitive to human visual system. As a result real edges in the blocks are preserved. Finally corner outlier is detected and removed using an algorithm. This paper focuses to remove the blocking artifacts while preserving the real edges. In this paper new evaluation parameter DMSD is proposed. Experimental results shows that this proposed method is an efficient one to remove the blocking artifacts & to improve the quality of the image.

Index Terms - BDCT, blocking artifacts, image compression , postprocessing, ,SingularValue Decomposition



1. INTRODUCTION

Image data compression is a very important issue for many applications in the field of visual communications. The objective of image compression is to reduce irrelevance and redundancy of the image data in order to be able to store or transmit data in an efficient form while maintaining image quality. Many efficient image coding techniques have been developed for various applications and among them DCT based compression has been in use over a long period of time as a international standard for image and video compression. In a typical DCT compression scheme, the input image is divided into small blocks (typically 8×8), each block being transformed independently to convert the image elements to DCT coefficients. The DCT coefficients are then quantized using a scalar quantizer & all of the quantized DCT coefficients are encoded using variable length encoding. At the decoder end, the received data is decoded, de-quantized, and reconstructed by the Inverse DCT (IDCT).

This block DCT scheme takes advantage of the local spatial correlation property of images and also saves processing time.

At high or moderate bit rates, the DCT coded image yields excellent reproduction without noticeable artifacts. However, at low bit rates, the reconstructed images generally suffer from visually annoying artifacts. One major drawback in blocking artifacts which is due to the loss of correlation between adjacent blocks. Blocking artifacts are introduced by the coarse quantization of transform coefficients at low bit rates and the independent quantization for each block. There are three kinds of blocking effects in JPEG decompressed images. One is the staircase noise along the image edges, another is the grid noise in the monotone area, and the third one is the corner outlier in the corner point of the 8×8 DCT block. The proposed post processing algorithm, which consists of three stages, reduces these blocking artifacts efficiently.

Nonlinear space variant filter [2] which adapts to the varying shape of the local signal spectrum & reduces only the locally out of band noise. It does not guarantee the smoothness of pixel values at either side of an edge. Here the classification is based on available edge information extracted from received blocky image. Thus the classification is inaccurate & hence the performance of this filtering scheme degrades. In [3] post processor is a combination of linear & non linear filter. The adaptation is achieved by changing filter coefficients according to the local characteristics of images & the blocking effects. Drawback of this method is computational complexity & no filtering is done at edge pixels. This filter used in [4] makes use of transform coefficients of shifted blocks, rather than neighboring blocks. To characterize the block activity, human visual system(HVS) sensitivity at different frequencies is considered. This method is having high computational complexity. Since the filter is not shapeadaptive, grid noise cannot be smoothed out& the edge features got blurred.

- Shanty Chacko working as a Assistant Professor in Dept of Electronics and Communication Engineering in Karunya University, Coimbatore, India.
Email: shanty@rediffmail.com
- Greeshma K.S is currently pursuing masters degree program in communication systems from Karunya University, Coimbatore, India.
Email: ksgreeshma10@gmail.com
- Jayakumar J working as a Assistant Professor in Dept of Electrical and Electronics Engineering in Karunya University, Coimbatore, India.
Email: jayakumar@karunya.edu

Smoothing of areas with sharp transitions will not eliminate blockiness. In [5] DCT domain algorithm extracts all parameters needed to detect the presence and to estimate the blocking artifacts by exploiting the properties of human visual system(HVS).Block shifted domain filtering makes the process complex.

In [6] reduction of blocking artifacts in smooth regions of the image is taken into consideration. Here the correlation between the intensity values of the boundary pixels of two neighboring blocks in the DCT domain is used to distinguish between smooth and non-smooth regions. This technique preserves the edge information. In [7] a 2D multiple notch filter is used to reduce the blocking artifacts. Both nonlinear space-variant filtering and adaptive filtering schemes are used. It requires classification of image blocks, which is based on available edge information extracted from the received blocky image. Hence the performance degrades. A post processing algorithm to reduce the blocking artifacts in JPEG compressed images is given in [8] In this technique image is classified into edge area and monotone area according to the edge map. The signal adaptive filtering consists of a 1-D directional smoothing filtering for edge area and 2-D adaptive average filtering for monotone area. In [9] An over complete wavelet representation is used to reduce the quantization effects of block based DCT is given. Chou et al. remove blockiness by performing a simple nonlinear smoothing of pixels [10]. They first form the maximum likelihood estimation of quantization noise to differentiate between artificial and actual edges.

In this paper artifacts of reconstructed image is reduced using the following algorithms proposed .The quality of the image is measured using two parameters ie.PSNR & DMSD.The validity of the DMSD technique is discussed in section[IV].

In the section II.A, the non-linear, adaptive filter which is used in [1] is described in detail. In the second section SVD based algorithm in [1] is used with a modification. In the final section corner outlier detection & replacement filter in [11] is also used with a modification.

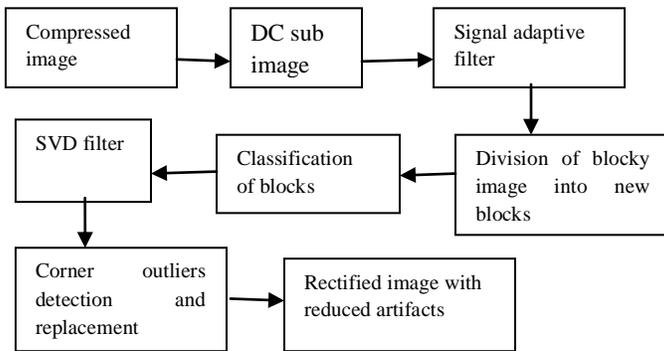


Fig : 1. Block diagram

2. PROPOSED METHOD

2.1.Post filtering of DC coefficients

In this method space variant, Non-linear, adaptive filter is used to reduce the discontinuities due to the DC coefficients. A signal adaptive filter is proposed to reduce both the grid noise in the relatively monotone area and the staircase noise along the image edge without any significant loss of the image details. In this Post filtering masking effect & adaptive filtering mechanisms are used.

2.1.1 Consideration of Masking Effect

For a good post processing method, it is desirable to efficiently reduce blocking artifacts and simultaneously to maintain image details. To measure the visibility of blocking artifacts, the masking phenomenon is considered here. Visibility of artifacts dependent on the local content of the given image. Due to the masking effect artifacts in regions of high activity are less perceptible than those in low-activity regions. The initial value of the mask at centre block of the image is taken as 0.75& the mask value of adjacent blocks of center block can be taken as 0.0625. The choice of a small neighborhood and a large weight set to the central component is to avoid blurring image details in those blocks of high activity These values which provides good results as compared to other .

2.1.2 Adaptive filtering mechanism

One of the factor of artifacts is due to the independent coding of blocks. So this filtering mechanism focuses on correlation among the blocks. Correlation within the blocks will differ depending upon whether it is monotone or edge. Classification of blocks can be done by comparing sum of absolute value of AC coefficients with a threshold. Threshold is average of AC coefficients. If the sum corresponding to center block is greater than threshold then the center block is considered as an edge block .In this case central will be given a weight factor 1 and surrounding blocks given a weight factor of zero otherwise center block is having the mask value 0.75 and surrounding blocks have 0.0625. If sum corresponding to the adjacent blocks is greater than threshold then it decreased to zero & the mask at the center block is increased by the same amount otherwise values remain unchanged. These values fed to the original & inverse DCT is taken.

2.2. Classification of blocks

This image is divided into the blocks in a manner in which the edges of the original image comes at the center of new blocks. Then the edges appears as high frequency components. This will help in reducing discontinuities due to these edges in the following SVD filter. Then the classification of blocks occurs which helps in processing blocks depending upon its activity. Blocking artifacts will appear more in low- activity blocks whereas it is less noticeable in high activity area.

2.3. Application of SVD filter

In this step block based, non-linear filter based on SVD [1] is used with a modification. One of the advantages of this technique is reduced computational complexity since the SVD is applied to each blocks & also it does not require any additional priori information. The singular value decomposition of an $m \times n$ real or complex matrix is of the form $M=U\Sigma V^*$, U is an $m \times m$ unitary matrix, Σ is an $m \times n$ diagonal matrix with non negative real numbers on the diagonal, V^* , an $n \times n$ unitary matrix. V^* which denotes the conjugate transpose of the matrix of V .

$$M=U\Sigma V^* = \sum_{i=0}^k \alpha_i u_i v_i^T \quad (1)$$

The diagonal entries of Σ ie $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \dots$ are known as the singular values. These values are arranged in descending order. The m columns of U and n columns of V are the left singular vectors of M & right singular vectors of M . The number of non-singular values which represents the rank r of the matrix. Here the singular values can be divided into significant & non-significant singular values depending upon the value of threshold. The threshold which is the 5% of the largest singular value.

In the paper [1], the singular values which are less than a threshold is reduced to zero. But in this method singular values less than threshold value is a fraction (k) of actual SVD value This value is varied from zero to one. Consider a $k \times k$ block in the image. Calculate the SVD of this block using the equation (1). The value of non-significant singular values which can be varied as

$$S(m,m)=k*s(m,m) \quad (2) \text{ where } k \text{ is a multiplication factor which is a fraction, } s(m,m) \text{ which indicates singular values.}$$

Calculate SVD for each block & reconstruct the image. The fraction at which best value of PSNR & DMSD (Discussed in section III) obtained is differing and it is varying with k for different images and is tabulated in table no:2

2.4 Corner outlier detection and reduction algorithm

Corner outlier is visually annoying artifact which appears in the form of stair case at the corner of the block. A corner outlier is characterized by a pixel which is much larger or much smaller than its neighboring pixels in the corner point of the 8×8 DCT block of the decompressed image.

In the filter [8], the pixel having corner outlier problem is considered to be only one third of the other adjacent corner pixels. Therefore when the pixel affected by corner outlier is much more than other adjacent corner pixels, this filter gives a distorted output. In the proposed paper, both these

possibilities are considered and modified the algorithm as follows

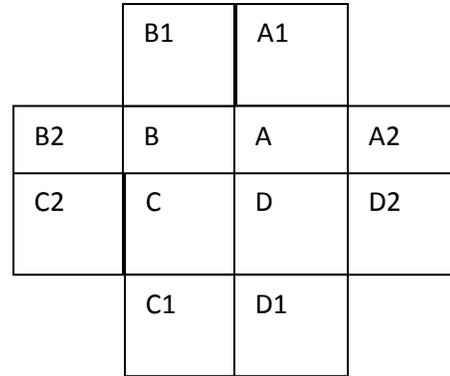


Fig : 2. Pixels used in corner outlier and reduction algorithms

Algorithm

Step 1:

if $abs(A-C) > abs(B-D)$ & if $abs(A-B) > abs(C-D)$ & if $abs(A-C) > m$ & $abs(A-B) > m$ & $abs(A-D) > m$ are satisfied, then $a = (A+B+C+D)/5$

If A is less than all other corner pixels adjacent to it. Then $a = (2*A+B+C+D)/5$
 Then A is replaced by a .

STEP 2:

Then $A1$ and $A2$ are replaced by $a1$ and $a2$ where $a1 = (2*A1+a)/3$; $a2 = (2*A2+a)/3$;

Similarly this algorithm has to be followed for remaining corner pixel points B, C and D .

3. EVALUATION PARAMETER DMSD

The objective quality of the restored image is usually evaluated by the peak signal-to-noise ratio (PSNR). PSNR alone does not indicate the quality of the image. So a new measuring parameter difference in mean square difference (DMSD) is introduced in this paper.

DMSD is given by,
 $DMSD = (MSD_o - MSD_r) / MSD_o \quad (3)$

$$MSD_o = \sum_{m=1}^r \sum_{n=1}^{c-1} [Io(m,n) - Io(m,n+1)]^2 - \sum_{m=1}^{r-1} \sum_{n=1}^c [Io(m,n) - Io(m+1,n)]^2 \quad \text{and}$$

$$MSD_r = \sum_{m=1}^r \sum_{n=1}^{c-1} [Ir(m,n) - Ir(m,n+1)]^2 - \sum_{m=1}^{r-1} \sum_{n=1}^c [Ir(m,n) - Ir(m+1,n)]^2$$

Where MSD_o denotes the mean square difference of the original image and MSD_r denotes the mean square difference of reconstructed image.

Five images are compressed at different compression rates using different quantization tables Q1,Q2,Q3,Q4 which given in the appendix.

From the PSNR and DMSD values tabulated in table no:1, it can be observed that as the quantization rate gets increased, the value of PSNR decreases & DMSD becomes more negative which indicates greater amount of artifacts. If the DMSD become more positive which indicates that the image is blurred. So image with high PSNR & with less negative DMSD indicates a high quality image. Figures 1(a),1(b),1(c) and 1(d) are the compressed images at various levels of quantization.

4. RESULTS & DISCUSSIONS

To evaluate the performance of this post processing algorithm, it has been applied to several jpeg compressed images. These images are compressed using DCT compression and encoded using non linear encoding. Various performance measures are used to evaluate the quality of the image. Two parameters used in this algorithm are PSNR & DMSD.

Peak signal to signal ratio(PSNR) is defined as

$$PSNR \text{ in db} = 10 \log_{10}(\max^2/MSE) = 20 \log_{10} (\max/MSE) \quad (4)$$

Where max is the peak signal for 8 bit PCM and its value is 255.

$$PSNR \text{ in db} = 20 \log_{10} (255/MSE)$$

MSE is mean square error given by

$$MSE = 1/N^2 [\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (rij - rij')^2] \quad (5)$$

r_{ij} and r'_{ij} are the pixel values of original and decoded images at position (i,j). PSNR alone does not indicate the quality of the image. Here PSNR with DMSD is used to indicate the quality of the image.

In table no:2, results shows that DMSD and PSNR is varying with different values of k. This shows these two parameters is a function of k. Fig 4, 5 and 6 indicates the processed images at different levels of algorithm. Fig:4 (a) is the decoded image using DCT after the process of compression. 4(b) indicates the image after DC component filtering. This method shows significant reduction in blocking artifacts. This can be observed in Figures 5(a),5(b) and 6(a),6(b). It can be verified by the calculation of PSNR and DMSD values tabulated in table no:3 and 4. The improvement in PSNR and DMSD shows the reduction of artifacts. After the application of adaptive filtering, an SVD based filter is further applied to reduce the blocking artifacts. This can be observed in figures 4(c),5(c) and 6(c). PSNR and DMSD values obtained through the application of this filter is tabulated in table no:3 and 4. In the final stage, corner outlier detection and reduction algorithm is applied. This algorithm reduces the artifacts and corresponding PSNR and DMSD values are plotted in the table. The corresponding figures are 4(d),5(d) and 6(d).

4.1 Comparison of the performance with other algorithm

In table no:5 PSNR and DMSD values of the proposed algorithm and reference [1] is tabulated. From the table, it can be observed that the PSNR and DMSD values are high as compared to the reference[1]. This shows that the proposed method is more efficient.



(a)



(b)



(c)



(d)

Fig 3: Zelda image compressed at various quantization levels. (a),(b),(c),(d) indicates image compressed at respective quantization levels Q1,Q2,Q3 and Q4

TABLE 1

PSNR AND DMSD VALUES OF COMPRESSED IMAGES AT VARIOUS LEVELS OF QUANTIZATION

Image	Q1		Q2		Q3		Q4	
	PSNR	DMSD	PSNR	DMSD	PSNR	DMSD	PSNR	DMSD
Lena	34.0367	-0.1104	24.3973	-1.3859	19.8312	-3.2404	15.6162	-5.2383
Zelda	31.5980	-0.5631	25.5933	-1.4220	20.9960	-3.1124	17.3681	-3.4805
Clown	29.3381	-0.2142	22.9506	-1.2096	18.8371	-2.3566	15.4684	-3.4720
Goldhill	26.7900	-0.5013	22.0619	-1.1685	18.9326	-1.5296	16.3168	-1.5754
Crowd	27.9558	-0.0133	21.7209	-0.8239	17.6137	-2.0641	14.3869	-2.8279
Elaine	31.0743	-0.0389	24.5906	-0.7857	20.9488	-1.4127	17.1600	-2.2004

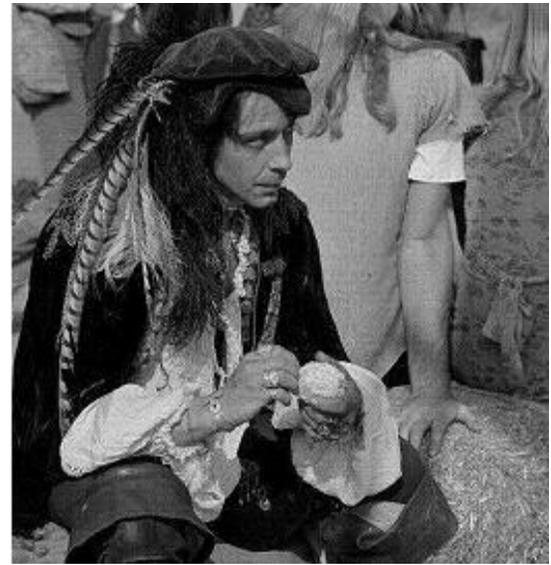
TABLE 2

VALUES OF PSNR AND DMSD WITH VARYING VALUES OF K

Image	k=0.1		K=0.3		K=0.5		K=0.8		k=0.9	
	PSNR	DMSD								
Lena	33.6162	0.1181	33.9117	0.1186	34.0367	0.1104	33.8774	0.0819	33.7370	0.0680
Barbara	24.9167	-0.3774	25.0068	-0.3826	25.0171	-0.3954	24.8827	-0.4287	24.8002	-0.4435
Bridge	22.3199	-0.9015	22.3974	-0.9193	22.3663	-0.9634	22.1231	-1.0785	21.9949	-1.1300



(a)



(b)



(c)



(d)

Fig 4: Result Man image at various levels of algorithm.(a) DCT compressed. (b) DC component filtering (c) SVD based filtering (d) With corner outliers reduction



(a)



(b)



(c)



(d)

Fig 5:Result Cameraman image at various levels of algorithm.(a) DCT compressed. (b) DC component filtering. (c) SVD based filtering (d) With corner outliers reduction



(a)



(b)



(c)



(d)

Fig 6: Result Bird image at various levels of algorithm.(a) DCT compressed. (b) DC component filtering.(c) SVD based filtering (d) With corner outliers reduction

TABLE 3

PERFORMANCE OF THIS ALGORITHM AT VARIOUS STAGES : **PSNR** VALUES OBTAINED USING DIFFERENT COMPRESSED IMAGES.

Image	Quantization Level	DCT coded	DC Component filtering	SVD based filter	Corner outliers reduction
		PSNR	PSNR	PSNR	PSNR
man	Q1	24.9550	24.9229	25.3386	25.3390
	Q2	18.7817	18.7820	19.0554	19.0580
	Q3	16.6836	16.7017	16.8282	16.8347
	Q4	14.2501	14.3010	14.3285	14.3454
cameraman	Q1	32.3170	32.2974	32.4565	32.3577
	Q2	22.3661	22.3762	22.6082	22.6051
	Q3	19.1067	19.1225	19.2255	19.2286
	Q4	15.6506	15.6860	15.7249	15.7487
Bird	Q1	23.3356	23.3289	23.6863	23.7014
	Q2	19.7434	19.7478	19.9997	20.0011
	Q3	16.8838	16.8946	17.0247	17.0316
	Q4	14.0819	14.0583	14.0906	14.1157

TABLE 4

PERFORMANCE OF THIS ALGORITHM AT VARIOUS STAGES : **DMSD** VALUES OBTAINED USING DIFFERENT COMPRESSED IMAGES

Image	Quantization level	DCT coded	DC Component filtering	SVD based filter	Corner outliers reduction
		DMSD	DMSD	DMSD	DMSD
Man	Q1	-0.4991	-0.4981	-0.3394	-0.3392
	Q2	-1.4181	-1.4171	-1.2893	-1.2856
	Q3	-1.2624	-1.2571	-1.1946	-1.1837
	Q4	-1.1988	-1.1809	-1.1658	-1.1425
Cameraman	Q1	0.3268	0.3279	0.3452	0.3349
	Q2	-0.9802	-0.9789	-0.8960	-0.8934
	Q3	-0.9558	-0.9536	-0.9143	-0.9068
	Q4	-1.5694	-1.5589	-1.5377	-1.5015
Bird	Q1	-0.8642	-0.8640	-0.7108	-0.7033
	Q2	-0.9752	-0.9747	-0.8727	-0.8718
	Q3	-1.3759	-1.3741	-1.3051	-1.2550
	Q4	-1.5195	-1.5046	-1.4769	-1.4383

TABLE 5
 PERFORMANCE COMPARISON OF PROPOSED METHOD WITH REFERENCE [1]

Image	Quantization level	Proposed method		Reference[1]	
		PSNR	DMSD	PSNR	DMSD
Man	Q1	25.3390	-0.3392	24.9815	-0.4781
	Q2	19.0580	-1.2856	18.8011	-1.2901
	Q3	16.8347	-1.1837	16.7056	-1.2364
	Q4	14.3454	-1.1425	14,3012	-1.2022
Cameraman	Q1	32.3577	0.3349	32.2823	0.3794
	Q2	22.6051	-0.8934	22.4042	-0.9946
	Q3	19.2286	-0.9068	19.1283	-1.1032
	Q4	15.7487	-1.5015	15.6867	-1.7104
Bird	Q1	23.7014	-0.7033	23.3763	-0.8879
	Q2	20.0011	-0.8718	19.7596	-1.0538
	Q3	17.0316	-1.2550	16.8982	-1.2863
	Q4	14.1157	-1.4383	14.0586	-1.6879

5. CONCLUSION

In this paper a non-linear filter is used to reduce the difference in DC coefficients caused by compression. Then a SVD based approach is used to reduce blocking artifacts. The singular values less than threshold is replaced by a fraction of the original value. For various fractional values results are computed and compared. Also Corner outlier is detected and reduced using a filter. It is observed that overall quality of the compressed image is improved using these three techniques

QUANTIZATION TABLES

Q1

6	4	4	6	10	16	20	24
5	5	6	8	10	23	24	22
6	5	6	10	16	23	28	22
6	7	9	12	20	35	32	25
7	9	15	22	27	44	41	31
10	14	22	26	32	42	45	37
20	26	31	35	41	48	48	40
29	37	38	39	45	40	41	40

Q2

20	24	28	32	36	80	98	144
24	24	28	34	52	70	128	184
28	28	32	48	74	114	156	190
32	34	48	58	112	128	174	196
36	52	74	112	136	162	206	224

APPENDIX

80 70 114 128 162 208 242 200
 98 128 156 174 206 242 240 206
 144 184 190 196 224 200 206 208

Q3

50 60 70 70 90 120 255 255
 60 60 70 96 130 255 255 255
 70 70 80 120 200 255 255 255
 70 96 120 145 255 255 255 255
 90 130 200 255 255 255 255 255
 120 255 255 255 255 255 255 255
 255 255 255 255 255 255 255 255
 255 255 255 255 255 255 255 255

Q4

110 130 150 192 255 255 255 255
 130 150 192 255 255 255 255 255
 150 192 255 255 255 255 255 255
 192 255 255 255 255 255 255 255
 255 255 255 255 255 255 255 255
 255 255 255 255 255 255 255 255
 255 255 255 255 255 255 255 255
 255 255 255 255 255 255 255 255

REFERENCES

- [1] V.K. Srivastava, "A SVD based scheme for post processing of dct coded images", *Electronic Letters on Computer Vision and Image Analysis* vol. 8(3): pp.1- 14, 2009
- [2]. B. Ramamurthi, A. Gersho, "Nonlinear space-variant post-processing of block coded images", *IEEE Trans. Acoust. Speech Signal Process.* ASSP-34(5): 1258-1268, 1986.
- [3] C.J. Kuo and R.J. Hsieh, "Adaptive post-processor for block encoded images", *IEEE Trans. Circuit Syst.Video Technol.*,vol. 5 (4):pp. 298-304, August 1995.
- [4] T. Chen, H. R. Wu, and B. Qiu, "Adaptive postfiltering of transform coefficients for the reduction of blocking artifacts", *IEEE Trans. Circuits Syst. Video Technol*,vol. 11(5):pp. 594-602, May 2001.
- [5] S. Liu, A.C. Bovik, "Efficient DCT-domain blind measurement and reduction of blocking artifacts", *IEEE Trans. Circuits Syst. Video Technol.* Vol(12):pp. 1139-1149, 2002.

- [6] Y. Luo, R.K. Ward, "Removing the blocking artifacts of block-based DCT compressed images", *IEEE Trans. Image Process.* Vol.12 (7):pp. 838-842, July 2003.
- [7] V.K. Srivastava and G.C. Ray, "Design of 2D-multiple Notch Filter and Its Application in Reducing Blocking Artifact from DCT Coded Image", in Proc. 22nd Ann. *IEEE Int. Conf. Engineering in Medicine and Biology, Chicago 2000*: pp.2829-2833, July 2000.
- [8] Y.L. Lee, H.C. Kim, and H.W. Park, "Blocking effect reduction of JPEG images by signal adaptive filtering", *IEEE Trans. on Image Processing*, 7 (2): 229-234, February 1998.
- [9] Z. Xiong, M. T. Orchard, and Y.Q. Zhang, "A Deblocking Algorithm for JPEG Compressed Images Using Overcomplete Wavelet Representations", *IEEE transactions on circuits and systems for video technology*, VOL. 7(2), APRIL 1997
- [10] J.Chou, M. Crouse and K. Ramchandran, "A Simple Algorithm for Removing Blocking Artifacts in Block-Transform Coded Images", *IEEE Signal processing letters*, VOL.5(2), FEBRUARY 1998
- [11] J.A.Gandam, J.S. Sidhu "A Post-Processing Algorithm for Detection & Removal of Corner Outlier", *International Journal of Computer Applications (0975 - 8887)* Volume 4(2), July 2010
- [12] Sangjin Kim, Sinyoung Jun, Eunsung Lee, Jeongho Shin, and Joonki Paik, "Ringing Artifact Removal in Digital Restored Images Using Multi-Resolution Edge Map", *International Journal of Signal Processing, Image Processing and Pattern Recognition*, Vol. 2(4), December, 2009
- [13] Dung T. Vo, Truong Q. Nguyen, "Localized filtering for artifact removal in compressed images", In proc. International conference on Acoustics, speech and signal processing", May 2011
- [14] Dung T. Vo, Truong Q. Nguyen and Sehoon Yea, Anthony Vetro, "Adaptive Fuzzy Filtering for Artifact Reduction in Compressed Images and Videos", *IEEE transactions on Image processing*, Vol. 18, no. 6, June 2009.

Author's Biography



Mr. Shanty Chacko received the B.E from Manipal Institute of Technology, ME from Government College of Technology, Coimbatore. He is currently working as an Assistant Professor in Karunya University, Coimbatore. His research interests include Image processing and signal processing.



Ms. Greeshma K.S received B.E from Nehru Institute of Engineering and Technology, Tamilnadu. she is currently doing MTech from Karunya University, Coimbatore.



Mr. Jayakumar J received the M.E from Tyagaraja College of Engineering, Madurai, Tamilnadu. He received PHD from Anna University, Chennai

Artifacts Removal and Edge Detection of Digitally Compressed Images

Shanty Chacko, Lijo T Joseph, J Jayakumar

Abstract — Block discrete cosine transform compressed image exhibits visually annoying compression artifacts. In this paper a method for reducing the compression artifacts and to increase the visual quality of image is been presented. The compressed image is filtered using Gaussian filter to reduce the amount of artifacts. In order to reproduce the high frequency coefficients a differential image is been obtained. From the differential image the artifact contents are removed and smoothed edge parts are retained. The edge detected image is added with the filtered image to improve the sharpness of the image. Results show that the image thus restored achieves perceivable image quality.

Index Terms — Block discrete cosine transform (BDCT), compression artifacts, image restoration, JPEG, post-processing

1 INTRODUCTION

Image compression schemes are extremely used now a day's in order to accommodate with the bandwidth of storage medias. To better compress the image, block discrete cosine transform is a method that is widely used. First the image is divided into 8×8 nonoverlapping blocks and each block is transformed independently to convert image into dct coefficients followed by quantization and variable length encoding. Thus binary data streams are generated for data transmission. The BDCT is the recommended transform technique for both still and moving image coding standards, such as JPEG, H.261.H.263 and MPEG. Block discrete cosine transform along with the application of dc quantization give rise to discontinuities between the blocks termed as blocking artifacts. This is due to the fact that the two low frequency dct coefficients in the adjacent blocks similar in value gets quantized to different quantization bins. Removing the high frequency components result in ringing artifacts around the strong edges. Thus the image quality gets decreased and to improve the quality of images image restoration schemes are been used.

There are various approaches to suppress the artifacts in transform domain and spatial domain. Zeng [1] models the blocking artifact as 2-D step functions and reduces it by applying zero masking to the dct coefficients of some shifted image blocks. A signal adaptive filtering scheme is used to reduce blocking artifacts in [2] by means of an adaptive weighting mechanism and quantization. Here local region filtering of each pixel can be performed without the consideration of the predicted coordinates of block-iness and size

of the smoothing window can be varied.

- Shanty chacko is an assistant professor at the Department of Electronics and Communication Engineering in Karunya University, Coimbatore, India E-mail: shantyc@rediffmail.com
- Lijo T Joseph is currently pursuing masters degree program at the Department of Electronics and Communication Engineering in Karunya University, Coimbatore, India E-mail: lijo.joseph8@gmail.com

- J Jayakumar is an assistant professor at the Department of Electrical and Electronics Engineering in Karunya University, Coimbatore, India E-mail: jayakumar@karunya.edu

Certain dct coefficients are modified in [3] by changing the step edge into slope to reduce discontinuities between the pixels in the blocks. [4] proposed a method to reduce compression artifacts by utilizing the wavelet coefficients and by analyzing the statistical characteristics of block discontinuities. Image deblocking algorithms using overcomplete wavelet representation enables multiscale edge analysis and makes suitable for edge based image compression. In [5], blocking artifacts are removed in wavelet domain by exploiting cross-scale correlations among wavelet coefficients and extracting and protecting the edge information. Also the noise components get smoothed out. In [6] the high frequency coefficients are scaled by a factor that depends on the compression ratio and subtracted from the observed image which is then used to design an adaptive filter that depends on the statistical behavior of the preprocessed image.

In [7],[8] filtering is performed in shifted windows by the method of thresholding the neighborhood and applying quantization constraints. [9] presents a method to reduce discontinuities by local ac coefficient regularization of shifted blocks in the discrete cosine transform domain, block-wise shape adaptive filtering in the spatial domain, and quantization constraint in the DCT domain. Iterative postprocessing techniques, based on the theory of POCS are also been proposed. In [10] the deblocking algorithm presents the iterative procedure based on quantization constraint sets and smoothness constraint sets which are the most commonly used convex sets to restore the original image. The images degraded by noise are restored iteratively by using a priori information about the original image, and it deblocks images by iterating the projections onto the quantization constraint set and image smoothness constraint set until convergence. [11] deals with POCS method

by adjusting the pixel intensity and by human visual system modeling adjustment. Thus, if we view the blocking artifact as the noises, then we can apply the image restoration techniques to alleviate the artifact.

The rest of the paper is organized as follows. Section II describes in detail the proposed scheme for reducing artifacts by filtering and edge detection methods and establishing the validation criteria of the parameter difference in mean square difference (DMSD). In Section III, the simulation results are discussed. Finally, Section IV presents the concluding remarks.

2 PROPOSED METHOD

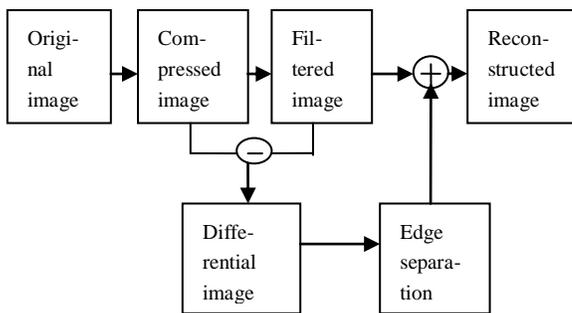


Fig.1 Block diagram of proposed scheme

2.1 Local filtering using Gaussian method

As a result of compression some artifacts are induced due to signal attenuation and noise. So postprocessing techniques are employed to reduce the discontinuities between various pixels in the image. The simplest approach among the postprocessing techniques is to filter the image. The blocking artifact is the result of an artificial boundary discontinuity between two adjacent blocks and ringing artifacts can be characterized by high-frequency components. Various filters are used to smoothen the image including Gaussian filtering, anisotropic filtering, neighborhood filtering, wiener filter or by means of threshold applications [12]. In this paper, local filtering is employed to degrade the image by removing the noise and attenuated coefficients. Although the attenuation intensity varies over different image regions, it has been assumed that the neighboring attenuation appears to be almost similar. The Euclidean distance is used to measure the neighborhood similarity of intensities to obtain the weights of individual pixels, which can be defined as,

$$w(i,j) = \sum_{m,n \in w_n} [(e^{-(I(m,n)-I(i,j))^2/h}) \times I(i,j)] / z(m,n) \quad (1)$$

where $I(m,n)$ is the value of neighboring pixels, $I(i,j)$ is the intensity level of central pixel and h is the degree of filtering control parameter which is to be set to a suitable value that depends upon the amount of discontinuities in the image. The filtering parameter decays weights as a Euclidean

distance function. $z(m)$ is the normalizing constant.

$$z(m,n) = \sum_{m,n \in w_n} (e^{-(I(m,n)-I(i,j))^2/h}) \quad (2)$$

The image resolution of the filtered image is finite and for different local regions the attenuation intensity will vary. Therefore in this paper a 5×5 window centered at pixel m (w_n) is employed for local filtering[13]. The filtered value of the central pixel thus becomes the weighted mean of the all pixels in its neighborhood. The filtering technique smoothes out the block boundary and reduces artifacts. But there is an expense of possible loss in high-frequency components, resulting in blurring of the decoded image.

2.2 Edge Detection

By subtracting the local filtered image from the compressed image, a differential image is obtained. This differential image mainly comprises of the high frequency components. A threshold is applied to detect sharp edges that are present in the differential image. The threshold is selected so as to give more edge information and to get an image which is less susceptible to artifacts. Coefficients are forced to zero for a value less than the threshold to reduce artifacts in the smoothened areas. The isolated non zero coefficients are also forced to zero since it does not possess any edge information. Finally the high frequency image is added with the filtered image [13] to enhance the sharpness of the image with reduced artifacts.

2.3 Evaluation of DMSD

The objective quality of the restored image is usually evaluated by the peak signal-to-noise ratio (PSNR). The higher the PSNR, the smaller is the difference between the restored image and the original image and the restored image is considered to be better. But higher PSNR value does not always implies that the quality of the image is good. The image with artifacts after compression will have higher PSNR value compared to the filtered image with lesser artifacts. With artifacts, the PSNR value is high since the all the pixel values are not changed from its original values. The above fact is clear in table 3 and 4. In order to validate the DMSD, seven images are compressed at different rates using the quantization table given in the appendix. The PSNR and DMSD calculated in each case is tabulated in table 1 and 2. From the quantization table shown in the appendix, it can be observed that as the quantization rate gets increased, the value of PSNR gets reduced due to greater amount of artifacts. From the table, it can be inferred that PSNR alone does not indicate the quality of the image. So a new measuring parameter called difference in mean square difference (DMSD) was also used to measure the quality of the image.

DMSD is given by,

$$DMSD = (MSD_o - MSD_r) / MSD_o \text{ where}$$

$$MSD_o = \sum_{m=1}^r \sum_{n=1}^c [Io(m,n) - Io(m,n+1)]^2 - \sum_{m=1}^{r-1} \sum_{n=1}^c [Io(m,n) - Io(m+1,n)]^2 \text{ and} \quad (3)$$

$$MSD_r = \sum_{m=1}^r \sum_{n=1}^c [Ir(m,n) - Ir(m,n+1)]^2 - \sum_{m=1}^{r-1} \sum_{n=1}^c [Ir(m,n) - Ir(m+1,n)]^2$$

MSD_o denotes the mean square difference of the original image and MSD_r denotes the mean square difference of reconstructed image. From the table 1, it can be observed that as the compression rate is increased DMSD, becomes more negative. The more negative value of DMSD indicates more amount of artifacts present in the image. So DMSD along with PSNR was used to measure image quality.

TABLE 1
EVALUATION OF PSNR FOR VARIOUS IMAGES COMPRESSED USING QUANTIZATION TABLE

Image	Q1	Q2	Q3	Q4
Lena	33.6407	24.0736	19.6928	15.5063
Flower	32.6035	25.5545	21.3246	17.3218
Moon	28.1283	23.5699	20.8169	18.0414
Peppers	28.3694	23.9335	19.7652	15.9101
Liftingbody	33.3299	27.4806	22.8664	17.4449
Zelda	31.0778	25.3600	20.8447	17.1571
Sailboat	26.4179	21.5348	17.8147	14.2483

TABLE 2
EVALUATION OF DMSD FOR VARIOUS IMAGES COMPRESSED USING QUANTIZATION TABLE

Image	Q1	Q2	Q3	Q4
Lena	-0.5408	-3.2531	-6.3206	-9.5218
Flower	-1.1588	-3.7477	-7.0178	-8.7492
Moon	-2.4233	-3.4273	-3.7472	-3.7743
Peppers	-1.3515	-1.8610	-3.5005	-4.7476
Liftingbody	-0.4705	-1.9204	-3.4889	-4.9010
Zelda	-1.5121	-3.0738	-5.9165	-7.0998
Sailboat	-1.5620	-2.7958	-4.5934	-5.3886

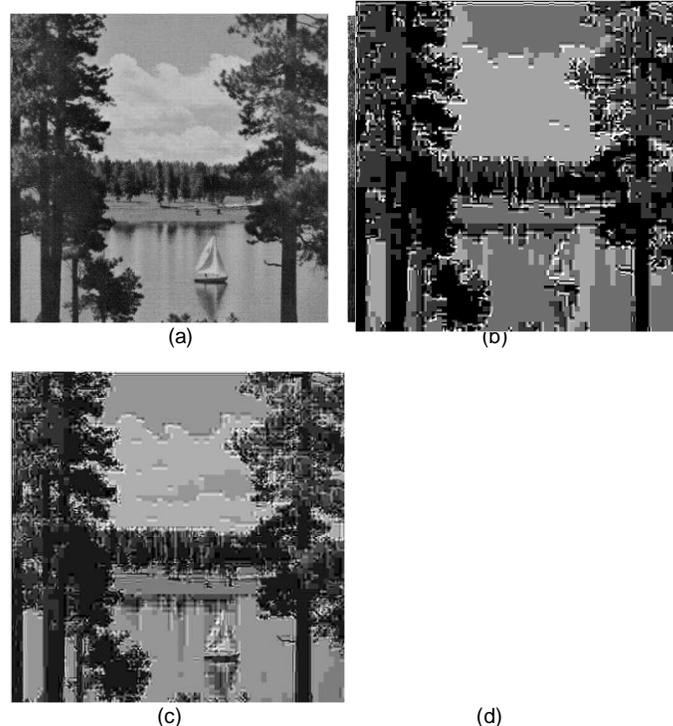


Fig. 3 Sailboat image compressed using different quantization tables (a) Q1 (b) Q2 (c) Q3 (d) Q4

3 RESULTS AND DISCUSSION

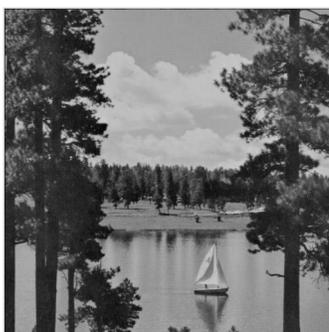
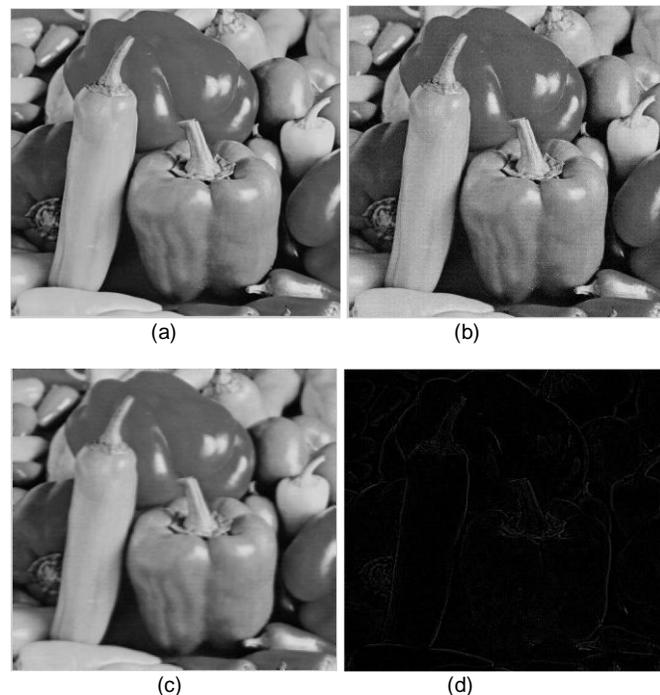


Fig. 2 Original Sailboat image

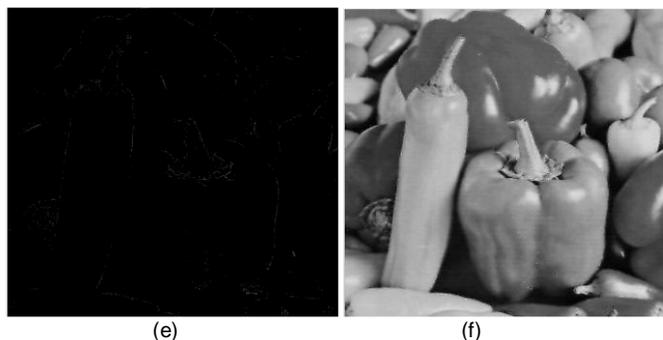


Fig. 4 Results of the proposed scheme (a) original image of Peppers (b) compressed image (c) filtered image (d) differential image (e) edge detected image (f) restored image

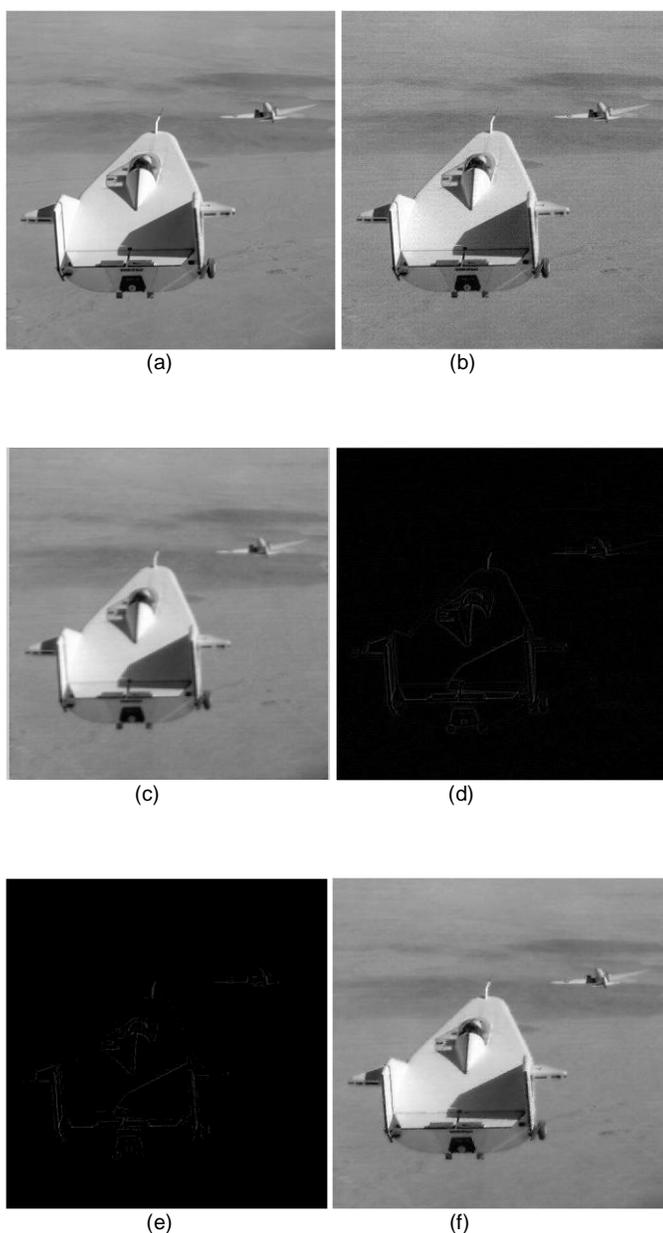


Fig. 5 Results of the proposed scheme (a) original image of Liftingbody (b) compressed image (c) filtered image (d) differential image (e) edge detected image (f) restored image

TABLE3
DMSD VALUES OF DIFFERENT IMAGES AT VARIOUS STAGES

	Compressed image	Filtered image	Reconstructed image
Lena	-0.5408	0.6690	-0.0942
Flower	-1.1588	0.7164	-0.0133
Moon	-2.4233	0.8325	-0.0003
Peppers	-1.3515	0.7891	0.0268
Zelda	-1.5121	0.7828	-0.0340
Sailboat	-1.5620	0.8316	-0.0068
Liftingbody	-1.1590	0.7951	-0.0222

TABLE4
PSNR VALUES OF IMAGES AT VARIOUS STAGES

	Compressed Image	Filtered image	Reconstructed image
Lena	33.6407	31.8299	33.6397
Flower	32.6035	33.2409	33.4624
Moon	28.1283	29.9279	30.1678
Peppers	28.3694	29.1036	29.3830
Zelda	31.0788	32.5250	32.5320
Sailboat	26.4179	26.5877	26.8416
Liftingbody	31.7837	31.3311	32.5555

The proposed method of restoration is tested on various images compressed by block discrete cosine transform. Images of size 512×512 have been used for testing. In this method, the degree of filtering is been set to 600. The value of threshold is been set as the 20% of the maximum absolute value of the coefficients in the differential image. The simulation has been done with the help of Matlab 7.9 software.

The PSNR and DMSD are the two parameters used to evaluate the quality of the reconstructed image. Comparing with the compressed image, the processed image shows lesser amount of artifacts. The sharpness of the image is also preserved since the high frequency coefficients are been added to the filtered image. Assessing only by using PSNR does not always give consistent results. From the results it was observed that for a filtered image, the PSNR gets reduced and the value of DMSD becomes more positive. The positive value of DMSD indicates that artifacts are been reduced. But the image gets smoothened. For increasing the quality of the restored image, the high frequency coefficients are been added to the filtered image. Thus the sharpness of the image is increased and only small numbers of DCT coefficients are modified. A visual assessment becomes highly necessary when compression is heavy. By

adjusting the DMSD and by obtaining a sufficient PSNR value an image is been restored with better visual quality.

4 CONCLUSION

This paper presented a method for improving the visual quality of block discrete cosine transform based compressed images. The high frequency coefficients obtained by thresholding differential image are added with filtered image to increase the image details. This method was developed to reduce unwanted signal disturbances while preserving the image sharpness and quality. The image quality is also been assessed in terms of PSNR and DMSD. The artifacts are significantly reduced and the overall visual appearance is greatly improved.

APPENDIX

QUANTIZATION TABLES

Q1

6	4	4	6	10	16	20	24
5	5	6	8	10	23	24	22
6	5	6	10	16	23	28	22
6	7	9	12	20	35	32	25
7	9	15	22	27	44	41	31
10	14	22	26	32	42	45	37
20	26	31	35	41	48	48	40
29	37	38	39	45	40	41	40

Q2

20	24	28	32	36	80	98	144
24	24	28	34	52	70	128	184
28	28	32	48	74	114	156	190
32	34	48	58	112	128	174	196
36	52	74	112	136	162	206	224
80	70	114	128	162	208	242	200
98	128	156	174	206	242	240	206
144	184	190	196	224	200	206	208

Q3

50	60	70	70	90	120	255	255
60	60	70	96	130	255	255	255
70	70	80	120	200	255	255	255
70	96	120	145	255	255	255	255
90	130	200	255	255	255	255	255
120	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255

Q4

110	130	150	192	255	255	255	255
-----	-----	-----	-----	-----	-----	-----	-----

130	150	192	255	255	255	255	255
150	192	255	255	255	255	255	255
192	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255

REFERENCES

- [1] B. Zeng, "Reduction of blocking effect in DCT-coded images using zero-masking techniques", *Signal Process.*, vol.79, pp.205-211, Dec. 1999.
- [2] T. Chen, H.R. Wu, and B. Qiu, "Adaptive postfiltering of transform coefficients for the reduction of blocking artifacts", *IEEE Trans. Circuits Syst. Video Technol.*, vol.11, pp.594-602, May. 2001.
- [3] S. Liu, and A.C. Bovik, "Efficient DCT-domain blind measurement and reduction of blocking artifacts", *IEEE Trans. Circuits Syst. Video Technol.*, vol.12, pp.1139-1149, Dec. 2002.
- [4] A.W.C. Liew, and H. Yan, "Blocking artifacts suppression in block-coded images using overcomplete wavelet representation", *IEEE Trans. Circuits Syst. Video Technol.*, vol.14, pp. 433-437, Apr. 2004.
- [5] Z. Xiong, M.T. Orchard, and Y.Q. Zhang, "A deblocking algorithm for JPEG compressed images using overcomplete wavelet representations", *IEEE Trans. Circuits Syst. Video Technol.*, vol.14, pp. 433-437, Apr. 1997.
- [6] A.S. Al-Fohoum, and A.M. Reza, "Combined edge crispiness and statistical differencing for deblocking JPEG compressed images", *IEEE Trans. Image Process.*, vol.10, pp. 1288-1298, Sep. 2001.
- [7] G. Zhai, W. Zhang, X. Yang, W. Lin, and Y. Wu, "Efficient image deblocking based on postfiltering in shifted windows", *IEEE Trans. Circuits Syst. Video Technol.*, vol.18, pp.122-126, Jan. 2008.
- [8] G. Zhai, J. Cai, W. Lin, X. Yang, and W. Zhang, "Shifted window based filtering for alleviating blocking artifacts", in *Proc. IEEE Workshop on Signal Processing Systems.*, pp. 289-294, Oct. 2008.
- [9] G. Zhai, W. Zhang, X. Yang, W. Lin, and Y. Wu, "Efficient deblocking with coefficient regularization, shape-adaptive filtering, and quantization constraint", *IEEE Trans. Multimedia.*, vol.10, pp.735-745, Aug. 2008.
- [10] A. Zakhor, "Iterative procedure for reduction of blocking effects in transform image coding", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, pp. 91-95, Mar. 1992.
- [11] J.J. Zou, and H. Yan, "A deblocking method for BDCT compressed images based on adaptive projections", *IEEE Trans. Circuits Syst. Video Technol.*, vol.15, pp. 430-435, Mar. 2005.
- [12] A. Buades, B. Coll, and J.M. Morel, "A non local algorithm for image denoising", in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition.*, vol.2, pp.60-65, Jun. 2005.
- [13] Lin Ma, Debin Zhao, and Wen Gao, "Learning based image restoration for compressed images", *Signal processing: Image communication.*, vol.27, pp.54-65, Jan. 2012

OPTIMUM RANGES FOR DATA TRANSMISSION IN MOBILE COMMUNICATIONS

Fazlay Rabby Reza

Abstract—Mobile communications are the fastest growing communications technology in history. In the field of data transmission, Optimum data transmission is a key factor for mobile communication which can be easily defined from Throughput, means the number of bits per second correctly received. Mobile communication allows transmission of voice and multimedia data via a mobile device using wired and wireless links; here these data are represented by bits. The throughput of a data communications system depends on a number of variables including packet size, transmission rate, the number of overhead bits in each packet, received signal power, received noise power spectral density, Signal to noise ratio. The purpose of this work is to find out the optimum ranges by observing certain changing parameter such as transmission rate, packet length, signal-to-noise ratio (SNR) with throughput depending on the very nature of the channels(AWGN & Rayleigh fading Channel). Here I ignore Rician fading channel because this occurs one of the paths, typically a line of sight signal, is much stronger so error occurring possibilities are less than others.

Index Terms— Optimum, Throughput, AWGN, Rayleigh channel, Packet length, Transmission rate, Signal-to-Noise Ratio (SNR).

◆

1 Introduction

In data transmission, throughput is the amount

of data moved successfully from one place to another in a given time period. The throughput is usually measured in bits per second (bit/s or bps), and sometimes in data packets per second or data packets per time slot. Typically, throughputs are measured in kbps, Mbps and Gbps, the speed with which data can be transmitted from one device to another. Data rates are often measured in megabits (million bits) or megabytes (million bytes) per second. These are usually abbreviated as Mbps and MBps, respectively. Throughput is affected by the channel environment such as the distance between the transmitter and the receiver, the fading state of the channel, the noise and interference power characteristics, the packet size, the transmission rate, the number of overhead bits in each packet, the modulation technique, and the channel conditions. From these variables, we can calculate other important quantities such as the signal-to-noise ratio, the binary error rate, and the packet success rate. [1]

In AWGN channel the thermal noise of the communication channels, the noise is additive means the received signal equals the transmit signal plus some noise, where the noise is statistically independent of the signal. The noise is white means the power spectral density is flat, so the autocorrelation of the noise in time domain is zero for any non-zero time offset and the noise samples have a Gaussian distribution. Mostly it is also assumed that the channel is Linear and Time Invariant. The most basic results further assume that it is also frequency non-selective.

In Rayleigh Fading Channel assume that the magnitude of a signal that has passed through such a transmission medium will vary randomly, or fade, according to a Rayleigh distribution and the delays associated with different signal paths in a multipath fading channel change in an unpredictable manner and can only be characterized statistically. When there are a large number of paths, the central limit theorem can be applied to model the time-variant impulse response of the channel as a complex-valued Gaussian random process.

Fazlay Rabby Reza has completed his graduation in Electronics & Telecommunication Engineering from Daffodil International University, Bangladesh in 2011. (e-mail: fazlayrabby.reza@gmail.com)

Here for finding optimum ranges for data transmission using throughput optimization for fixed packet length varying the value of the transmission rate, and for fixed transmission rate varying the value of the packet length and also throughput optimization as a function of SNR where both packet length & transmission rate varied. All these conditions are applied on two channels: Additive white Gaussian noise (AWGN) and Rayleigh fading channel.

2 Introductions To Throughput

2.1 Definitions

System throughput (T) is the number of payload bits which means the essential data that is being carried within a packet or transmission unit per second received correctly. If the payload transmission rate is (K/R) and the packet success rate $f(\gamma)$ which defined as the probability of receiving a packet correctly .so it can be written in the terms of bits per second,

$$T = \frac{K}{L} R f(\gamma) \quad (2.1)$$

This probability is a function of the signal-to-noise ratio which can be defined as

$$\gamma = \frac{E_b}{N_0} = \frac{P}{N_0 R} \quad (2.2)$$

Here $E_b = P/R$ joules is the received energy per bit. We will look at maximizing the throughput in a Gaussian white noise channel and Rayleigh Fading Channel with respect to the transmission rate and packet length.

2.2 Throughput Analysis

The transmitter, operating at R b/s, sends data in packets. Each packet contains L bits including a payload of K bits and a cyclic redundancy check error-detecting code with C bits. A forward error correction encoder produces the remaining L-K-C bits in each packet. The channel adds white noise with power spectral density N_0 watts/Hz and the

signal arrives at the receiver at a power level of P watts. In this research paper I assume to be the sum of all noise and interference, which can be modeled as Gaussian white noise. The CRC decoder detects transmission errors and generates acknowledgments that cause packets with errors to be retransmitted. Table 1 displays a summary of the variables in our analysis and their unit, notation. [2]

Table 1: Variables in Analysis

Quantity	Notation	Unit	Value
Signal to Noise Ratio	γ	dB	10/Varied with Transmission Rate
Received signal power	P	Watts	5×10^{-9} W
Receiver noise power spectral density	N_0	W/Hz	10^{-15} W/Hz
Binary transmission rate	R	bits/s	Varied
Packet size	L	bits	Varied
Cyclic Redundancy Check	C	bits	16 bits

We have used some equations for optimizing throughput and simulated those in MATLAB and then observed the results in graphical representation in MATLAB window. By analyzing these graphical representations we will suggest optimal range for throughput in each channel on each condition.

2.3 Principles For This Work

This analysis includes the following simplifying Principles on which we analyze throughput & it's optimization:

••The CRC decoder detects all errors in the output of the FEC decoder.

••Transmission of acknowledgments from the receiver to the transmitter is error free and instantaneous.

••In the presence of errors, the system performs selective repeat ARQ retransmissions.

••The received signal power is P watts, either a constant or a random variable with a Rayleigh probability density function, representative of fading wireless channels.

3 Throughputs vs. Transmission Rate (Fixed Packet Length)

The AWGN channel model is used to approximate the way errors are introduced in a data stream when it is transmitted over a lossy medium. The model we may use in the Workshop is the Additive White Gaussian Noise channel (AWGN). This channel model is memory less, meaning that the distortion of one bit is independent of all other bits in the data stream. Here one noise is added with the original transmitted signal, called white noise.

Frequency flat, fast Rayleigh fading may be considered the most critical disturbance in a wireless communication system. In its most general form, it is modeled as a multiplicative time continuous random (zero mean complex Gaussian) distortion of the transmitted signal. In order to achieve an efficient communication here, each part of the communication link must be carefully

designed based on the properties of the time continuous channel.

3.1 Equation Analysis

To find the transmission rate, $R=R^*$ b/s, that maximizes the throughput, we differentiate Equation (2.1) with respect to R to obtain:

$$\frac{dT}{dR} = (K/L) f(\gamma) + (K/L) R \frac{df(\gamma)}{d\gamma} \frac{d\gamma}{dR}$$

$$= (K/L) \left(f(\gamma) + R \frac{df(\gamma)}{d\gamma} (-P/N_0 R^2) \right) \quad (3.1)$$

Next we set the derivative to zero:

$$f(\gamma) - (P/N_0 R) \frac{df(\gamma)}{d\gamma} = f(\gamma) \gamma \frac{df(\gamma)}{d\gamma} = 0 \quad (3.2)$$

$$f(\gamma) = \gamma \frac{df(\gamma)}{d\gamma} \quad (3.3)$$

We adopt the notation $\gamma = \gamma^*$ for a signal-to-noise ratio that satisfies Equation (3.3). The corresponding transmission rate is

$$R^* = \frac{P}{\gamma^* N_0} \quad (3.4)$$

A sufficient condition for a locally maximum throughput at $R=R^*$ is:

$$\left. \frac{d^2T}{dR^2} \right|_{R=R^*} < 0 \quad (3.5)$$

The solution to Equation (3.3), γ^* , is the key to maximizing the throughput of a packet data transmission. To operate with maximum throughput, the system should set the transmission rate to R^* in Equation (3.4). γ^* is a property of the function, $f(\gamma)$, which is the relationship between packet success rate and signal to interference ratio. This function is a property of the transmission system including the modem, codecs, receiver structure and antennas. Each system has its own ideal signal-to-noise ratio, γ^* . Depending on the channel quality, reflected in the ratio P/N_0 , the optimum transmission rate is R^* in Equation (3.4).

For non-coherent FSK in a white Gaussian noise channel, the probability of a bit error is given by:

$$P(\gamma) = \frac{1}{2} e^{-\frac{\gamma}{2}}$$

For a model that corresponds to mobile radio communications, we can perform the same analysis for a fast fading Rayleigh channel. For non-coherent FSK in a Rayleigh fading channel, the probability of a bit error is given by:

$$\bar{P}(G) = \frac{1}{2 + G}$$

Here the G is the Signal to noise ratio and $\bar{P}(G)$ is the bit error rate for Rayleigh fading channel.[3]

3.2 Simulation For AWGN Channel

Here, I have taken simulation for different values of packet length (L) where transmission rate varies

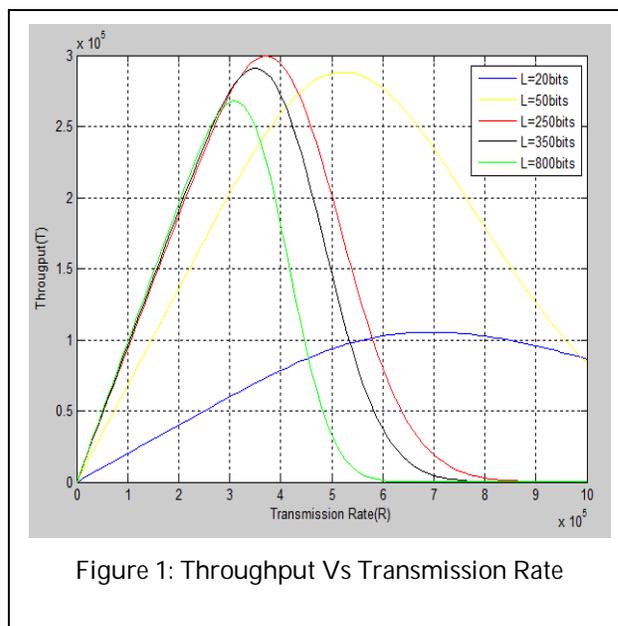


Figure 1: Throughput Vs Transmission Rate

with throughput.

In figure I have noticed that when the packet length size was small then the throughput has reached its highest peak with higher transmission rate and also has fallen in a wide range. But as soon as the packet length has kept higher, then the curve of throughput is steeper rather than flat and the throughput has reached its highest peak lower transmission rate. When we have increased our packet length size than before, then the throughput has reached the maximum peak at a very lower transmission rate and also has fallen down quite quickly.

3.3 Optimum Ranges

From the above analysis I have seen that if we keep our packet length less than 350 bits and greater

than 50 bits, then we will be able to get the maximum throughput and the transmission rate shouldn't be so high which is in a range of 300kbps to 600kbps. So using the general equations for calculating throughput in respect of transmission rate and keeping the packet length fixed the throughput can be optimized in a certain range.

3.4 Simulation For Rayleigh Fading Channel

Here, I have also taken simulation for different values of packet length (L) where transmission rate varies with throughput. This graph also contains five assumptions.

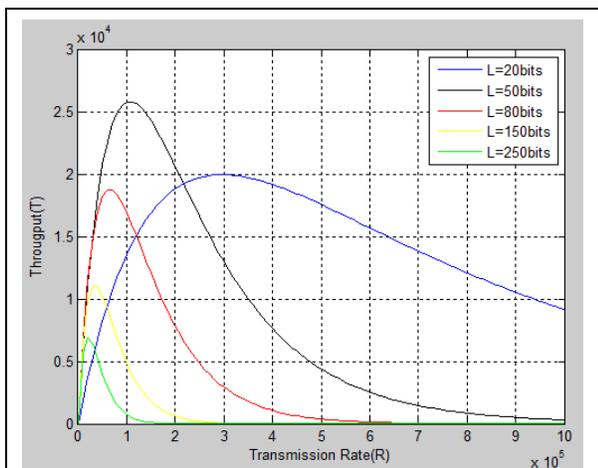


Figure 2: Throughput Vs Transmission Rate

In figure 2 I have noticed that this graph characteristic is behave almost like Figure 1. But here after certain packet length size, then the throughput has reached the maximum peak at a very lower transmission

3.5 Optimum Ranges

From the above representation of throughput with the function of transmission rate and fixed packet

length, I have also observed that for transmission rate of 50kbps to 450 Kbps we have got the highest peak of throughput and the packet size was within 20bits to 80bits.

3.6 IN COMBINED FORM

For showing the comparison between two channels for optimum ranges with the function of transmission rate where fixed packet lengths are fixed in same graph. we will choose four readings whose are exist in the optimum ranges.

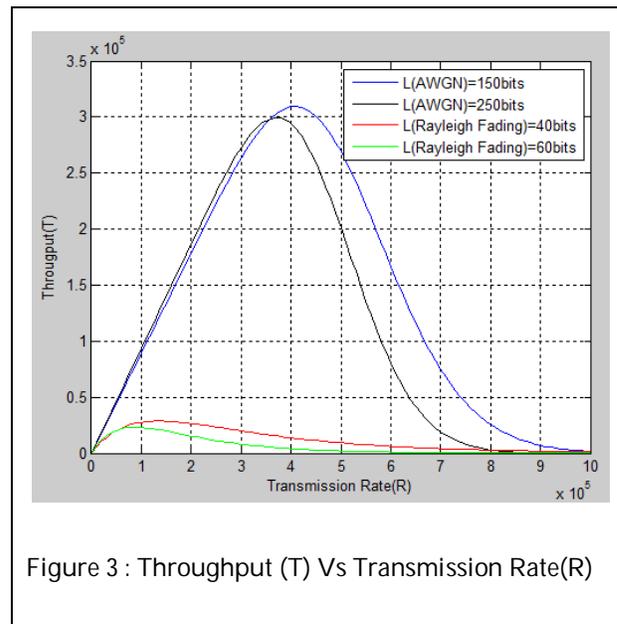


Figure 3 : Throughput (T) Vs Transmission Rate(R)

3.7 GRAPHICAL ANALYSIS

By observing the above figure I can say that in AWGN channel, data transmission ability is much more than Rayleigh Fading channel. But in AWGN channel, it takes more transmission rate to reach peak position than that of Rayleigh Fading channel.

4 Throughput (T) Vs Packet Length (L)

(Fixed Transmission Rate)

The AWGN channel models the distortion incurred by transmission over a lossy medium as the addition of a zero-mean Gaussian random value to each bit. Decoders can take advantage of the added information of "how close" a received value is to a valid bit value (0 or 1 for our purposes). This type of decoding is called soft decision decoding.

In the Rayleigh fading channel a higher diversity order is obtained if the encoder output bits, instead of symbols, are interleaved. By also using codes of lower rates, the diversity is increased even further. To avoid bandwidth expansion here, the channel symbol constellation must be expanded accordingly. The resulting system is referred to as bit-interleaved channel symbol expansion diversity (CSED) and is the best low-complexity coded modulation scheme presented so far for the Rayleigh fading channel.

4.1 Equation Analysis

Each packet, of length L bits, is a combination of a payload (K bits) and overhead (L-K bits). Because the packet success rate $f(\gamma)$ is a decreasing function of L, there is an optimum packet length, L^* . When $L < L^*$, excessive overhead in each packet limits the throughput. When $L > L^*$, packet errors limit the throughput. When there is no forward error correction coding, which we shall assume for the entirety of this chapter, ($K=L-C$, where C bits is

the length of the cyclic redundancy check), there is a simple derivation of L^* . In this case,

$$f(\gamma) = (1 - P_e(\gamma))^L, \quad (4.1)$$

Where $P_e(\gamma)$ is the binary error rate of the modem. Therefore, in a system without FEC, the throughput as a function of L is

$$T = \frac{L-C}{L} R (1 - P_e(\gamma))^L \quad (4.2)$$

To maximize T with respect to L, we consider L to be a continuous variable and differentiate Equation (4.2) to obtain

$$\begin{aligned} \frac{dT}{dL} &= R \frac{L-C}{L} (1 - P_e(\gamma))^L \ln(1 - P_e(\gamma)) \\ &+ R \frac{C}{L^2} (1 - P_e(\gamma))^L \quad (4.3) \end{aligned}$$

Setting the derivative equal to zero produces a quadratic equation in L with the positive

root:

$$L^* = \frac{1}{2}C + \frac{1}{2} \sqrt{C^2 - \frac{4C}{\ln(1 - P_e(\gamma))}} \quad (4.4)$$

As shown in Figure 4 and 5 (in which $C=16$), the optimum packet size is a decreasing function of $P_e(\gamma)$.

As the binary error rate goes to zero, the packet error rate also approaches zero and the optimum packet size increases without bound. Because

$P_e(\gamma)$ decreases with γ , L^* increases monotonically with signal-to-noise ratio. Better channels support longer packets. Of course, in practice L is an integer and the optimum number of bits in a packet is either the greatest integer less than L^* or the smallest integer greater than L^* .

Equations (3.3) and (4.4) can be viewed as a pair of simultaneous equations in variables L and γ . Their simultaneous solution produces the jointly optimum packet size and Signal - to- noise ratio of a particular transmission system. We will use the notation, L^{**} and γ^{**} , respectively for the jointly optimized variables.

For non-coherent FSK in a white Gaussian noise channel, the probability of a bit error is given by:

$$P(\gamma) = \frac{1}{2} e^{-\frac{\gamma}{2}}$$

For a model that corresponds to mobile radio communications, we can perform the same analysis for a fast fading Rayleigh channel. For non-coherent FSK in a Rayleigh fading channel, the probability of a bit error is given by:

$$\overline{P}(G) = \frac{1}{2+G}$$

Here the G is the Signal to noise ratio and $\overline{P}(G)$ is the bit error rate for Rayleigh fading channel.[3]

4.2 Simulation For AWGN Channel:

For throughput optimization for fixed transmission rate varying the value of the packet length we have taken graphs based on transmission rate..The

graph contains five assumptions where SNR will be changed with the value of R .

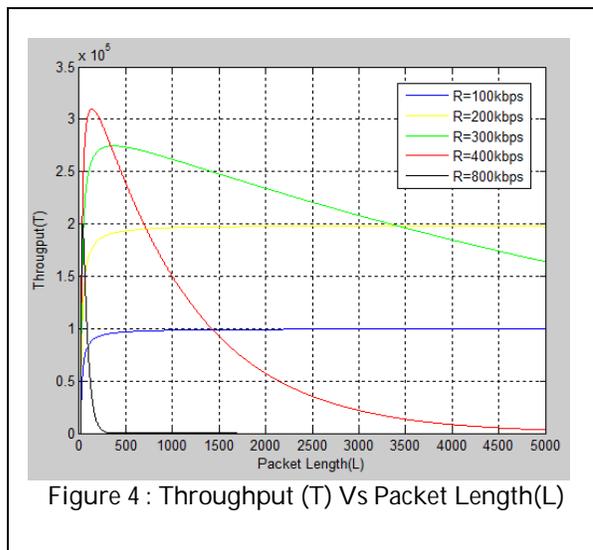


Figure 4 : Throughput (T) Vs Packet Length(L)

In the figure we see that first two readings throughputs are gone to the saturation after a certain packet length. Then two readings, peak throughputs are increased with the increase of transmission rate. But as soon as the transmission rate has kept higher, then the curve of throughput steeper rather than flat and the throughput has reached its highest with small packet size and has fallen down in a short range respectively.

4.3 Optimum Ranges

I have observed that if we keep our transmission rate in the range of 300kbps to 400kbps we will be able to get the maximum throughput. And for the maximum throughput the packet has come in the range of 50bits to 600 bits.

4.4 Simulation For Rayleigh Fading Channel

In the figure I have noticed that for the lower transmission rate the throughput has reached its highest peak with higher packet size and the highest peak which is fallen down in a wide range. The transmission rate has kept higher, then the curve of throughput is steeper than before. Then

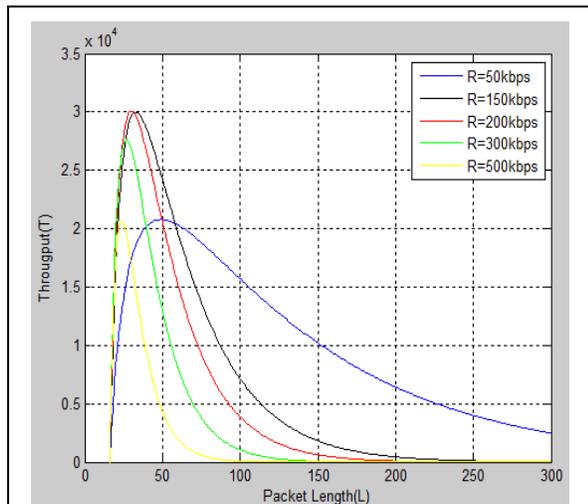


Figure 5: Throughput (T) Vs Packet Length (L)

more transmission rate less will be the throughput and fallen down in a quite short range.

4.5 Optimum Ranges

From these graphs I have observed that for transmission rate of 150 Kbps to 300 kbps we have got the optimum throughput. But here we get maximum throughput at very small packet length within 20bits to 50 bits.

4.6 IN COMBINED FORM

For showing the comparison between two channels for optimum ranges for this condition where transmission rates are fixed in same graph. we will choose four readings whose are exist in the optimum ranges.

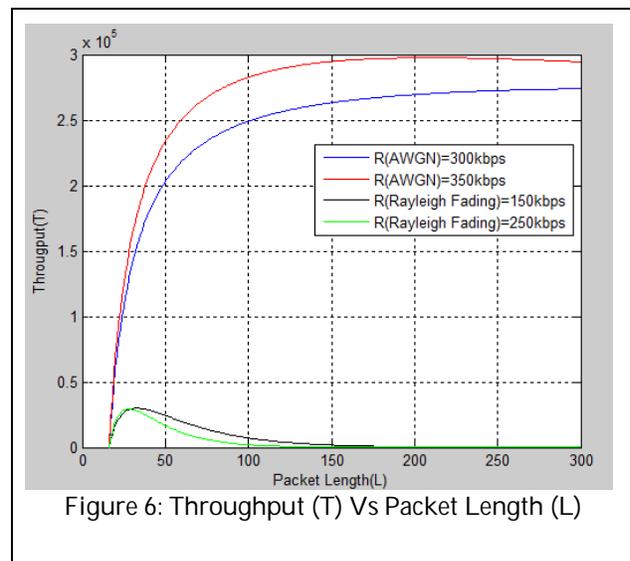


Figure 6: Throughput (T) Vs Packet Length (L)

4.7 GRAPHICAL ANALYSIS

I have taken four readings for comparison. For the higher transmission rate the throughput increases with the increase of packet length in AWGN channel but in the Rayleigh fading channel we get peak throughput with little packet length which has fallen in a short range. So, the graphical representation is as like as before (condition).

5 Throughput (T) Vs SNR

(Joint Optimization)

Some important conclusions can be drawn from the previous information for the channels. We first notice that at high SNR values (low transmission rates) the packet length used to maximize the throughput must be large. When the transmission rate increases and the SNR drops, the packet length to maximize the throughput must also decrease. Another observation we make is how the throughput curve behaves for increasing values of L when different SNR values are used. From previous figures we can see that at high bit rates (low SNR) the choice of packet size is more critical.

5.1 Joint Optimization

The following figure has shown the throughput optimization with respect to joint optimization in terms of SNR (Signal to Noise Ratio) in both channels. We have analyzed throughput with

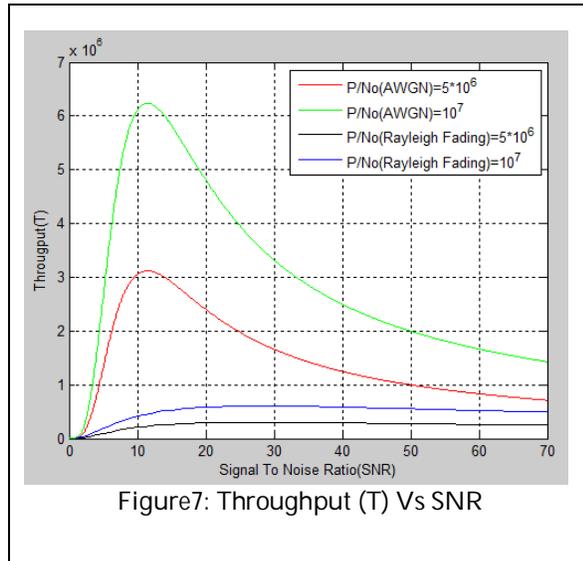


Figure7: Throughput (T) Vs SNR

respect to SNR and has kept some fixed values for P/N_0 . Here we have taken another value of $\frac{P}{N_0}$ that is 10^7 .

5.2 GRAPHICAL ANALYSIS

Observing the above figure I can say that throughput is maximized in this condition for both channels. But as like before throughput is maximized in AWGN channel. From the figure we

REFERENCES

[1] Richard O. LaMarire and Arvind Krishna, "Maximizing Throughput in a Random Access Radio System by Optimal Power Level Choice," Communications, 1996. ICC '96, Conference Record, Converging Technologies for Tomorrow's Applications. pp. 614 - 620, vol. 1, 1996.

see that the value of P/N_0 is changed which is doubled. Then in both channels the peak throughput is doubled from the peak value where P/N_0 is 5×10^6 .

5.3 Optimum Ranges

From this optimum SNR, the optimum packet length and (given the received signal power over the noise power spectral density) the optimum transmission rate can be determined.

6 Conclusions

We obtained some important information from this work. We have learned that data throughput depends on a wide variety of variables. Given a modulation scheme, and the channel conditions, the optimum SNR at which to operate the system for maximum throughput can be derived. We have talked about many different variables and how changing certain parameters can yield better throughput performance. Implementing these concepts in real systems is not as easy as one might think. In order for a system to make parameter changes it must be able to make simple measurements in the system.

[2] Peter E. Bausbacher and Jeffrey L. Kearns, "Transmission Parameter Selection in an Adaptive Packet-Radio Network," Tactical Communications Conference, 1990. Vol.1, pp. 51 - 68, 1990.
[3] Richard J. Lavery : Throughput Optimization for Wireless Data Transmission. *Master's Thesis Work*, Polytechnic University on Publication Date: June 2001.

Secured Communication through Fibonacci Numbers and Unicode Symbols

A. Joseph Raphael, Dr. V. Sundaram

Abstract— The objective of cryptography is to make it feasible for two persons to exchange a message in such a way that other persons cannot understand. There is no end to the number of ways this can be done, but here the proposed method will be more concerned with a technique of encoding the text in such a way that the recipient can only discover the original message. The original message usually called plain text is converted into cipher text by finding each character in the message and replacing it with another character based on the Fibonacci number generated. Further cipher text is converted into Unicode symbols, which avoid suspicion from the third party when sent through an unsecured communication channel. There are two levels in the proposed method; (i) converting plain text to cipher text and (ii) converting cipher text to Unicode symbols. In each level, security key is used to encode the original message which provides two levels of security from intruders. On the other end, the extraction algorithm is designed in such a way that the process converts the Unicode symbols into cipher text and then cipher text to plain text. This encoding and decoding scheme of the proposed method is significantly different as compared to the traditional methods.

Index Terms— Cipher text, Decryption, Encryption, Fibonacci Number, Key, Plain text, Unicode Symbols

1 INTRODUCTION

From the dawn of civilization to the highly networked societies that we live in today, communication has always been an integral part of our existence. What started as simple sign communication centuries ago have evolved into many forms of communication today, the internet being just one such example. Methods of communication today include radio communication, telephonic communication, network communication and mobile communication. All these methods and means of communication have played an important role in our lives, but in the past few years, network communication, especially over the internet, has emerged as one of the most powerful methods of communication with an overwhelming impact on our lives. Such rapid advances in communications technology have also given rise to security threats to individuals and organizations.

Cryptography is the art and science of secret writing. The term is derived from the Greek language *kryptos* means secret and *graphos* means writing. Encryption is the actual process of applying cryptography. Much of cryptography is math oriented and uses patterns and algorithms to encrypt messages, text, words, signals and other form of communication. Cryptography has many uses, especially in the areas of espionage, intelligence and military operations. Today, many security systems and companies use cryptography to transfer information over the Internet. Some of this encryption is highly advanced; however, even simple encryption techniques can help uphold the privacy of any person.

2 CRYPTOGRAPHY

The most ancient and basic problem of cryptography is secure communication over an insecure channel.

2.1 Related Terms

Cryptology [3] encompasses both cryptography and cryptanalysis and looks at mathematical problems that underlie them.

Cryptosystems are computer systems used to encrypt data for secure transmission and storage.

Plain text is a message or data which are in their normal, readable form.

Encryption is encoding the contents of the message in such a way that hides its contents from outsiders.

Cipher text results from plaintext by applying the encryption key.

Decryption is the process of retrieving the plaintext back from the cipher text.

Substitution cipher involves replacing an alphabet with another character of the same alphabet set.

Mono-alphabetic system uses a single alphabetic set for substitutions.

Poly-alphabetic system uses multiple alphabetic sets for substitutions.

Caesar cipher is a mono-alphabetic system in which each character is replaced by the third character in succession. Julius Caesar used this method of encryption.

A digital signature is a block of data that is generated by the sender of a message using his/her secret key.

2.2 Cryptographic Goals

Authentication: is the process of providing proof of identity of the sender to the recipient; so that the recipient can be assured that the person sending the information is who and what he or she claims to be [1].

Privacy/confidentiality: is the process of keeping information private and secret, so that only the intended recipient can understand the information.

Integrity: is the method to ensure that information is not tampered with during its transit or its storage on the network. Any unauthorized person should not be able to tamper with the information or change the information during transit.

Non-repudiation: is a mechanism to prove that the sender really sends this message.

2.3 Types of Cryptographic Algorithms

In general, there are three types of cryptographic schemes typically used to accomplish the cryptographic goals: secret key (or symmetric) cryptography, public-key (or asymmetric) cryptography, and hash functions. In all cases, the initial unencrypted data is referred to as *plaintext*. It is encrypted into *ciphertext*, which will in turn (usually) be decrypted into usable plaintext.

2.3.1 Secret key cryptography

The process of encryption and decryption of information by using a single key is known as single key cryptography or symmetric key cryptography. In symmetric key cryptography, the same key is used to encrypt as well as decrypt the data. The main problem with symmetric key algorithms is that the sender and the receiver had to agree on a common key.

2.3.2 Public key cryptography

This cryptography technique is based on a combination of two keys—secret keys and public key. It is also known as asymmetric encryption. In this process, one key is used for encryption, and the other key is used for decryption. This process is known as asymmetric cryptography because both the keys are required to complete the process, and these two keys are collectively known as the key pair. In asymmetric cryptography, one of the keys is freely distributable and this key is called the public key which is used for encryption. Hence, this method of encryption is also called public key encryption. The second key is the secret or private key and is used for decryption. The private key is not distributable, like its name suggests, is private for every communicating entity.

2.3.3 Hash Functions

Hash functions, also called message digests and one-way encryption, are algorithms that use no key. Instead, a fixed-length hash value is computed based upon the plaintext that makes it impossible for either the contents or length of the plaintext to be recovered. Hash algorithms are typically used to provide a *digital fingerprint* of a file's content often used to ensure that the file has not been altered by an intruder or virus. Hash functions are also commonly employed by many operating systems to encrypt passwords.

2.3.4 Drawbacks of Existing Methods

- Transmission time for documents encrypted using public

key cryptography is significantly slower than symmetric cryptography.

- The key sizes must be significantly larger than symmetric cryptography to achieve the same level of protection.
- Public key cryptography is susceptible to impersonation attacks.

3 UNICODE

3.1 Birth of Unicode

ASCII, a character set based on 7-bit integers is still popular and its provision for 128 characters was sufficient at the time of its birth in the 1960s. The growing popularity of personal computing all over the world made ASCII inadequate for people speaking and writing many different languages with different alphabets. Newer 8-bit character sets, such as the ISO-8859 family, could represent 256 characters. This solution was good enough for many practical uses, but was a bit limiting for all the languages in the world [4]. In the end, the other parts of the world began creating their own encoding schemes, and things started to get a little confusing. It became apparent that a new character encoding scheme was needed, and the Unicode standard was born.

3.2 What is Unicode?

Unicode is a character encoding standard that has widespread acceptance. Unicode defines a large number of characters and assigns each of them a unique number, the Unicode code, by which it can be referenced. This encoding standard provides the capacity to encode all the characters used for the written languages of the world. The objective of Unicode is to unify all the different encoding schemes so that the confusion between computers can be limited as much as possible. The most common Unicode encodings are called UTF-n, where UTF stands for Unicode Transformation Format and n is a number specifying the number of bits in a basic unit used by the encoding. Two very common encodings are UTF-16 and UTF-8. In UTF-16, which is used by modern Microsoft Windows systems, each character is represented as one or two 16-bit (two-byte) words provides code point for more than 65000 characters (65536). Unix-like operating systems, including Linux, use another encoding scheme, called UTF-8, where each Unicode character is represented as one or more bytes [4]. The benefit of Unicode is that, it assigns each character a unique value and symbol, no matter what the platform, no matter what the program, no matter what the language [5].

4 NEW APPROACH USING FIBONACCI NUMBERS AND UNICODE SYMBOLS

Cryptography has now become an industry standard for providing information security, trust, controlling access to resources, and electronic transactions. Its use is no longer limited to just securing sensitive military information. In the proposed method, the plain text is converted into cipher text

Characters: v w x y z a b c defghijklmnopqrstuvwxyz-
zabc...

Fibonacci : 1 2 3 5 8 ...

Cipher Text: v w x z c

The character set follows a round-robin method and the character which falls below the Fibonacci number will be taken as the character in the cipher text. If there are a number of characters in the plain text, the process of finding the replacing character with the Fibonacci number might be difficult because each time the size of the Fibonacci number is increased by adding the previous two numbers. If the application does not support the size of the Fibonacci number, after a fixed range, Fibonacci number can be restarted from the beginning. Since the selection of the character depends on the Fibonacci number, it provides more security for the system, and any unknown person cannot decode the message easily.

4.1.2 Cipher Text to Unicode symbols

In the second level of security, the ASCII code of each character obtained from the cipher text plus the ASCII code of its previous character, and next character is added to the ASCII code of the equivalent character in the original message. Here, ASCII codes of four characters are used as a security key to further encode the characters available in the cipher text to Unicode symbols. For instance,

Cipher Text: v w x z c

$$\begin{aligned}
 & \rightarrow 98(b) + 99(c) + 100(d) + 79(O) = 376 \\
 & \rightarrow 121(y) + 122(z) + 123() + 76(L) = 442 \\
 & \rightarrow 119(w) + 120(x) + 121(y) + 76(L) = 436 \\
 & \rightarrow 118(v) + 119(w) + 120(x) + 69(E) = 426 \\
 & \rightarrow 117(u) + 118(v) + 119(w) + 72(H) = 426
 \end{aligned}$$

The decimal numbers obtained are converted into hexadecimal values to find its equivalent Unicode symbols. These symbols are saved in a text file which can be sent to the recipient through an unsecured communication channel. By looking at the symbols in a text file no unknown persons can identify what it is and the message cannot be retrieved unless the retrieval procedure is known.

Steps involved in Encryption

1. A sender wants to send a Hello message to a recipient.
2. The original message, also called plaintext, is converted to cipher text by using a key and Fibonacci numbers. The algorithm being used can produce a different output each time it is used, based on the key selected.
3. Cipher Text is converted into Unicode symbols using another key and saved in a text file.
4. The text file is transmitted over the transmission medium.

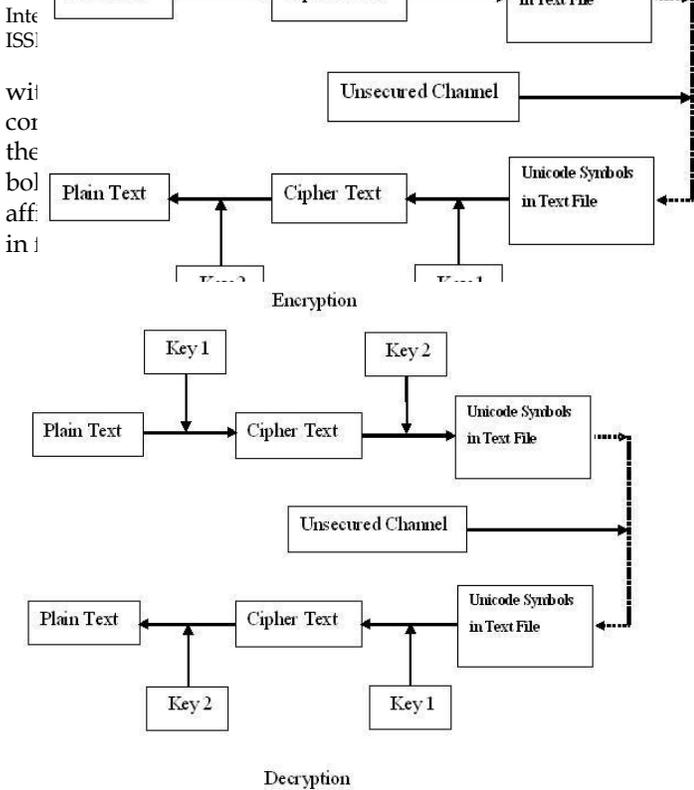


Figure 1: Affine Transformation of Encryption and Decryption Process

4.1 Encryption Method

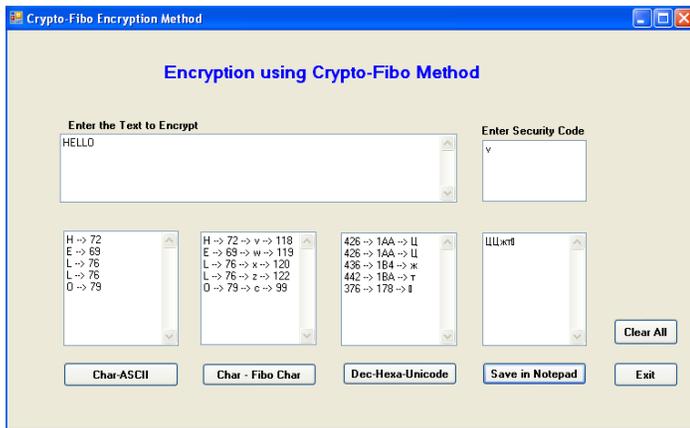
In the proposed method, the original message called plain text is converted into cipher text with the help of Fibonacci numbers generated. Here, each character is extracted from the original message and is replaced with another character, the way the character is chosen to replace the original character makes this method unique and different when compared to the traditional methods. The obtained cipher text is converted into Unicode symbols, and these symbols in a text file are transmitted to the recipient through an unsecured communication channel. Since the message is encrypted in two levels, it is hidden from others and makes the decryption process more difficult for any intruders. The conversion of plain text to Unicode symbols undergoes two phases namely; converting plain text to cipher text and converting cipher text to Unicode symbols.

4.1.1 Plain Text to Cipher Text

The conversion from the plain text to cipher text can be explained through an illustration. Let us consider a message to be encrypted and send through an unsecured channel as "HELLO." Each character is replaced with another character based on the Fibonacci number and security key chosen. Any one character is chosen as a first security key to generate cipher text. The characters in the cipher text depend on the security key chosen, and the Fibonacci numbers generated. For instance, let the first security key chosen be v.

Plain Text: H E L L O

4.1.3 Implementation



4.2 Decryption Method

The decryption process follows the reverse process of encryption with the help of two keys. At the recipient end, from the received text file each symbol is extracted and mapped to find its equivalent hexadecimal value, further the obtained value is converted into a decimal value to find out the plain text using the key. Without the knowledge of the key, an unknown person cannot even suspect the existence of any secret message in these decimal numbers.

Key (chosen to encrypt): v

Characters : v w x y z a b c efg-hijklmnopqrstuvwxyzabc...

Fibonacci Numbers : 1 2 3 5 8

Obtained Decimal Numbers: 426 426 436 442 376

To extract the original plain text, ASCII code of each character from the cipher text plus the ASCII code of its previous character and the next character is subtracted from each obtained decimal number. The remainder is the ASCII code of character in plain text; the accumulated characters form the original plain text.

$$426 - (117 (u) + 118 (v) + 119 (w)) = 72 (H)$$

$$426 - (118 (v) + 119 (w) + 120 (x)) = 69 (E)$$

$$436 - (119 (w) + 120 (x) + 121 (y)) = 76 (L)$$

$$442 - (121 (y) + 122 (z) + 123 (l)) = 76 (L)$$

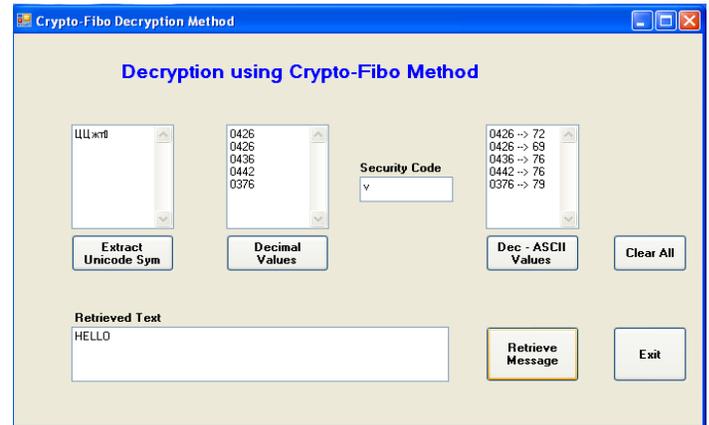
$$376 - (98 (b) + 99 (c) + 100 (d)) = 79 (O)$$

Steps involved in Decryption

1. At the recipient end, Unicode symbols are converted to hexadecimal values and then to equivalent decimal values using the same algorithm and key that were used to encrypt the message.
2. Perform subtraction with the decimal values and the ASCII code of characters from cipher text

3. The remainder after subtraction gives the ASCII code of the character in need.
4. The process is repeated for the number of characters in the cipher text and accumulated characters forms the original plain text.

4.2.1 Implementation



5 COMBINED CRYPTO-STEGANOGRAPHY

Steganography is not the same as cryptography. Data hiding techniques have been widely used for transmission of hiding secret messages for a long time. Ensuring data security is a big challenge for computer users. Business men, professionals, and home users all have some important data that they want to secure from others. Even though both methods provide security, to add multiple layers of security it is always a good practice to use Cryptography and Steganography together. By combining, the data is encrypted and the cipher text is embedded in Unicode symbols with the help of key. The combination of these two methods will enhance the security of the data embedded. This combined chemistry will satisfy the requirements for communication such as capacity, security and robustness to secure data transmission over an open channel.

6 ADVANTAGES

The proposed method is a process that scrambles information by rearrangement and substitution of content making it unreadable to anyone except the person capable of unscrambling it. Security key is provided in converting plain text to cipher text, and another key is used to convert the cipher text to Unicode symbols. Security key in each level is an added advantage to the method. It is difficult to decode the Unicode symbol from the text file which makes the system complicated in retrieval of the message for an unknown person. Moreover, information stored in a text file in the form of symbols increases the amount of information to be conveyed in secret. The advantage of using Unicode symbol reduces the suspicious while transmitting the file through an unsecured communica-

tion channel, else if any intruder tries to convert the symbol to the hexadecimal number nothing can be retrieved unless the retrieval process and the key is known.

7 CONCLUSION

Cryptography has evolved from an ancient science to an important area of research to secure communications. It has evolved from simple substitution ciphers to quantum cryptography. This method provides the means and methods of hiding data, establishing its authenticity, and preventing its undetected modification or unauthorized use. Furthermore, this presents a scheme that can transmit large quantities of secret information and provide secure communication between two parties. Any kind of text data can be employed as a secret message and is sent over the open channel, in the proposed procedure is simple and easy to implement.

REFERENCES

- [1] http://media.wiley.com/product_data/excerpt/94/07645487/0764548794.pdf, Chapter 1, Basics of Cryptography
- [2] <http://cs-exhibitions.uni-klu.ac.at/index.php>
- [3] I. Venkata Sai Manoj, "Cryptography and Steganography", International Journal of Computer Applications (0975 - 8887), Volume 1 - No.12
- [4] Michał Kosmulski, Introduction to Unicode, <http://www.linux.com/archive/articles/39911>
- [5] The Unicode Consortium, <http://www.unicode.org>
- [6] B. B. Zaidan, A.A. Zaidan, A.K. Al-Frajat and H.A. Jalab, "On the Differences between Hiding Information and Cryptography Techniques: An Overview", Journal of Applied Sciences (2010), ISSN 1812-5654
- [7] A. Joseph Raphael and Dr. V.Sundaram, "Secured Crypto-Stegano Communication through Unicode", World of Computer Science and Information Technology Journal Vol. 1, No. 4,138-143, 2011, ISSN: 2221-0741
- [8] Dipti Kapoor Sarmah, Neha bajpai, "Proposed System for Data Hiding Using Cryptography and Steganography", International Journal of Computer Applications (0975 - 8887), Volume 8 - No. 9, October 2010.
- [9] Sashikala Channalli and Ajay Jadhav, "Steganography An Art of Hiding Data", International Journal on Computer Science and Engineering Vol.1(3), 2009, 137-141.

AUTHORS PROFILE

A. Joseph Raphael obtained his Master degree in Computer Science from St. Joseph's College, Tiruchirapalli and Master of Philosophy from Alagappa University, Karaikudi. Currently, he is a PhD research scholar at Karpagam University, Coimbatore, India and also working as a lecturer in the department of Information Technology, Ibra College of Technology, Sultanate of Oman.

Dr. V. Sundaram earned his PhD in mathematics from Madras University. He is a research guide of Anna University, Coimbatore and Karpagam University in the field of computer

science and computer applications. He is currently guiding several PhD students in the areas of theoretical computer science, network security, cryptography and data mining. He has published several papers in national and international journals and organized 5 national conferences. He is a life member of ISTE and a member of Computer Society of India.

An approach for image noise identification using minimum distance classifier

Raina , Shamik Tiwari ,Deepa Kumari ,Deepika Gupta

Abstract :- This paper deals with the problem of identifying the nature of noise in order to apply the most appropriate algorithm for de-noising. The key idea involves isolation of some representative noise samples and extraction of their features for noise identification. The isolation of the noise samples is achieved through application of filters. Statistical features are extracted and the minimum distance classifier is applied for identification of the noise type present.

Keywords :- Noise, noise identification, statistical features, minimum distance classifier.

◆

1. INTRODUCTION

The identification of the nature of the noise affecting an image is an important stage in all information interpretation systems by vision when the nature of the degradation is unknown. The majority of filtering algorithms (Lee, Kuan, ...) [1] [2] and certain algorithms of contour detection (Canny, Deriche, ...) [3] [4] found in literature, assume that the nature of the noise and its statistical parameters are known. Whereas in most practical cases we have not a priori knowledge on these data [5]. For this reason the statistical parameters of the noise must be estimated as they condition the quality of the filtering or the analysis of the images [6]. In [7], we proved that it is possible to identify the nature of the noise by recording variations of local statistics (the standard deviation as a function of the average) computed in the homogeneous regions of the observed image. If the recording is parallel to the average axis, then the noise is declared as an additive one and its standard deviation is equal to the sampling average of the different values of the local standard deviation. If the recording can be assimilated by a line passing through zero, then the noise is declared as a multiplicative one and its standard deviation is given by the slope of the line. And finally, if the recording can not be viewed as a line passing through zero, then the noise is declared as an impulsive one. The previous methods presented in [7] [8] [9] are based on the criterion of maximum likelihood for the selection of the most homogeneous masks (Lee, Nagao etc.), from which the value of the local standard deviations are calculated. However, the disadvantage with this approach is the estimation of parameters from pixels belonging to masks a priori decided. This means that the estimates of standard deviations are sometimes necessarily biased and the final identification rates inevitably decreased in the case of

images degraded either by a weak multiplicative or an impulsive noise.

The search for efficient image de-noising methods is still a valid challenge at the crossing of functional analysis and statistics. In spite of the sophistication of the recently proposed methods, most algorithms have not yet attained a desirable level of applicability. All show an outstanding performance when the image model corresponds to the algorithm assumptions but fail in general and create artifacts or remove image fine structures

In order to increase the rate of identification and to improve the estimation of statistical noise parameters, we propose a new method. The statistical parameters kurtosis and skewness are calculated and the Minimum Distance Classifier is applied. Classification includes a broad range of decision-theoretic approaches to the identification of images. All classification algorithms are based on the assumption that the image in question depicts one or more feature and that

each of these features belongs to one of several distinct and exclusive classes. Classification analyzes the numerical properties of various image features and organizes data into categories.

2. THE PROPOSED METHOD

In principle, the noise identification method proposed here consists of three key steps:

- Step 1. Extract some representative noise samples from the given noisy image,
- Step 2. Estimate some of their statistical features, and
- Step 3. Use a simple pattern classifier to identify the type of noise.

In this paper, we consider four different types of commonly occurring image noise, namely,

uniform white, Gaussian white, speckle, and salt-and-pepper noise. Among these four types, speckle noise is of multiplicative type, whereas the other three are additive in nature. The filters selected for the above four types of noise are: Wiener filter [10] for uniform or Gaussian white noise, Homomorphic filter [11] for speckle noise, and median filter [10],[11] for salt-and-pepper noise. Also, the statistical

features studied here include "kurtosis" and "skewness". Table 1 lists the probability density functions, "kurtosis" and "skewness" values, and the selected filters for the four types of noise. From Table 1, we can see that different type of noise has different kurtosis or skewness values and those differences can be used to identify the noise type.

Table 1
 PDF, Kurtosis,Skewness and Filter selected for four type of noises.

	Probability Density Function	Kurtosis	Skewness	Selected Filter
Non-Gaussian White	$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$	1.8	0	Wiener Filter
Gaussian White	$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	3	0	Wiener Filter
Speckle (exponential distribution)	$f(x) = \begin{cases} \frac{1}{\sigma^2} \exp(-\frac{x}{\sigma^2}), & x \geq 0 \\ 0, & x < 0 \end{cases}$	9	2	Homomorphic filter
Salt-and-Pepper	$f(x) = \begin{cases} p_a, & \text{for } x = \text{salt} \\ p_b, & \text{for } x = \text{pepper} \\ 0, & \text{otherwise} \end{cases}$	Large (typically greater than 20 or so, depending on the noise density)	0	Median Filter

3. IDENTIFYING THE NATURE OF NOISE

3.1 The Algorithm

1. Fetch the input images.
2. Introduce noise(through imnoise or rand).
3. Filter the image and hence obtain the noise sample.
4. Extract the features from the noise sample.
5. Apply the Minimum Distance Classifier and classify noise

3.2 The statistical features

3.2.1 Kurtosis

Kurtosis is any measure of the "peakedness" of the probability distribution of a real-valued random variable. It is a descriptor of the shape of a probability distribution and there are different ways of quantifying it for a theoretical distribution and corresponding ways of estimating it from a sample from a population.

3.2.2 Skewness

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable. The skewness value can be positive or negative, or even undefined. Skewness has benefits in many areas. Many models assume normal distribution; i.e., data are symmetric about the mean. The normal distribution has a skewness of zero. But in reality, data points may not be perfectly

symmetric. So, an understanding of the skewness of the dataset indicates whether deviations from the mean are going to be positive or negative.

3.3 Evaluation of the features

The evaluation of features is carried out using Matlab simulations. The functions kurtosis(X) and skewness(X) are used to calculate the respective values for the feature kurtosis and skewness for the image sample X. A range of feature values are calculated for each image sample and then the mean value is calculated using the mean(x1, x2, x3...) function which returns the mean value of input specified.

3.4 Application of Minimum Distance Classifier

This is applied so as to find out the minimum distance & classify as to which class the image belongs to. The distances are calculated on the values of the features being extracted as before i.e kurtosis and skewness. Euclidean distance is calculated by the formula:

$$d_i = \sqrt{(K_i - K_m)^2 + (S_i - S_m)^2}$$

For every image the distances are calculated for every mean value of the features. The minimum the distance to a class, the image belongs to that category of noise. The distance calculated thus helps to lead to a conclusion.

3.4 Experimental Results

The algorithm is applied on a set of images with Uniform, Gaussian, Speckle and Salt & Pepper Noise. The following confusion matrix shows the result:

Table 2
Confusion Matrix for Minimum Distance Classifier

	Uniform	Gaussian	Speckle	Salt & Pepper
Uniform	86.2	10	3.8	0
Gaussian	6.4	90	3	0.6
Speckle	0	2.3	90.19	7.5
Salt & Pepper	1.5	5	3.5	90

4.CONCLUSION

A simple technique for identifying the type of noise present in a noisy image is proposed in this paper. The proposed technique is quite general in nature and can be used with a variety of de-noising filters. The results of simulation studies seem to indicate that the method is capable of accurately determining the type of noise.

REFERENCES

- [1] J.S LEE. "Digital image enhancement and noise filtering by use of local statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-2, vol 11: pages 165-168, March 1980.*
- [2] D.T. KUAN. "Adaptive noise smoothing filter for images with signal dependent noise," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-7 vol 11: pages 165-177, March 1985.*
- [3] J. CANNY. "A computational approach to edge detection," *IEEE Transactions on Pattern Analyzers and Machine Intelligence, PAMI-7 vol 11: pages 679-698, 1986.*
- [4] R. DERICHE. "Using canny's criteria to derive an optimal edge detector recursively implemented," *International Journal on Computer Vision, vol 1: pages*
- [5] D. KUNDUR and D. HATZINAKOS. "Blind image deconvolution," *IEEE Signal Processing Magazine, 13(3): pages 43-64, May 1996.*
- [6] C. SPINU, C. GARBAY, and J.M. CHASSERY. "Une approche cooperative et adaptative pour la segmentation d'images," *Proceedings GRETSI-95, pages 609-612, Sept 1995.*
- [7] K. CHEHDI and M. SABRI "A new approach to identify the nature of the noise affecting an image," *Proceedings IEEE ICASSP 9.2, vol 111: pages 285-288, March. 1992.*
- [8] K. CHEHDI. "Automatic identification of noises for an optimal filtering," *Proceedings IEEE CCSP 93, pages 474477, March 1993.*
- [9] L. BEAUREPAIRE and K. CHEHDI. "Identification of degraded image by a multiplicative or additive noise," *Proceedings EUSIPCO-96, vol 11: pages 899-902, Sept. 1996*
- [10] J. S. Lim, *Two Dimensional Signal and Image Processing, Prentice Hall, 1990*
- [11] A. K. Jain, *Fundamentals of Digital Image Processing, Pentice Hall, NJ, 1989.*
- [12] M. C. Motwani, M. C. Gadiya, R. C. Motwani, and F. C. Harris, Jr., *Survey of image denoising techniques, Proceedings of GSPx 2004, Santa Clara, CA, September 27-30, 2004*
- [13] L. Beaurepaire, K. Chehdi, and B. Vozel, *Identification of the nature of noise and estimation of its statistical parameters by analysis*

of local histograms, Proc. 1997 IEEE International Conference on Acoustics, *Speech and Signal Processing, Vol. 4, pp. 2805-2808.*

[14] B. Vozel, K. Chehdi, L. Klaine, V. V. Lukin, and s. k. Abramov, Noise Identification and Estimation of its Statistical Parameters by Using Unsupervised Variational Classification, *Proc. 2006 IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 2, pp. 841-844.*

[15] G. Torricelli, F. Argeni, and L. Alparone, Modelling and assessment of signal-dependent noise for image de-noising, *Proceedings of EUSIPCO 2002, Vol. III, pp. 287-290.*

[16] Shamik Tiwari, Ajay Kumar Singh, Statistical Moments based Noise Classification using Feed Forward Back Propagation Neural Network, *Proc. 2011 International Journal of Computer Applications, Vol. 18, pp. 36-40*

[17] Tan, Shan; Jiao, Licheng , A unified iterative denoising algorithm based on natural image statistical models: derivation and examples, *Optics Express, vol. 16, issue 2, p. 975, 2008*

[18] Lakhwinder Kaur , Savita Gupta , R. C. Chauhan, Image Denoising using Wavelet Thresholding, Indian Conference On Computer Vision, *Graphics and Image Processing, Ahmedabad ,2002.*

[19] R.K.Kulkarni, S.Meher, Mrs. J.M.Nair , An Algorithm for Image Denoising by Robust Estimator, *European Journal of Scientific Research, ISSN 1450-216X Vol.39 No.3 (2010), pp.372-380 © EuroJournals Publishing, Inc. 2010*

[20] K. Chehdi, M. Sabri, A New Approach to identify the nature of noise affecting an image, *IEEE, 1992*

[21] Lionel Beaurepaire, Kacem Chehdi and Benoit Voel, Identification of the nature of noise and estimation of its statistical parameters by analysis of local histograms, *IEEE, 1999*

[22] B. Vozel, K. Chehdi, L. Klaine, Vladimir V. Lukin, Sergey K. Abramov, Imation of its statistical parameters by using unsupervised variational classification, *IEEE, 2006*

[23] T. Santhanam and S. Radhika, A Novel Approach to Classify Noises in Images Using Artificial Neural Network, *Journal of Computer Science 6 (5): 541-545, 2010 ISSN 1549-3636 © 2010 Science Publications.*

Home Automation Indian Scenario: A survey of Architectures and Technologies

Nagaraj Shet M.N. and Shreesha Chokkadi

Abstract— Home automation is the automatically controlling the day today activities of the home. Home automation involves the controlling of various devices of the home, taking care of the well being of the occupants, providing security to the occupants, monitoring health of the occupants. Home automation network involves wireless embedded sensors and actuators which monitor various devices used for home management. One of the most critical aspects in home automation is communication and network technology. In this article a discussion is done about the Indian needs, challenges faced as per the Indian requirements and shortage of resources in India. The already available home automation technology can be tuned to be cheaper and widely acceptable even at the remote areas.

Index Terms— Home automation, Home security, Energy management.

1 INTRODUCTION

WE are living in the era of minute by minute developments in new technologies: The demand of easy way of the life is the talk of the day. Engineering industries are focusing on the projects which facilitate their customers with comfortable and secure living. During the last five years there was an enormous growth in field of communication. Speed with which new things are happening in the mobile communication and its popularity to every household makes a man rethink about having such technological changes in other fields like energy management, home automation and security trying to merge it with a hand held device like mobile phone

The rapid and magnanimous growth of communication facilities and living standards, there is great demand for enormous quantity of energy and automation. At the same time conventional sources of energy are rapidly depleting and also at the same time the cost of energy is rising. As the energy saved is less expensive than to generate equivalent amount of energy it is seriously required to find out ways to judiciously save energy and at the same time provide sufficient energy to the consumers to fully enjoy the comforts provided by the technological development. Then only a sustainable economical growth is possible. If we don't plan properly definitely there will be serious energy crisis. To overcome this crisis the only alternative in sight is exploration of non-conventional sources of energy. As the renewable energy tapping is costly, to make it more economical and cost effective it is required to make a system with value added services like home automation and home security. We can also have distributed energy generation systems wherever possible at the place of consumption. This will reduce the transportation costs. Such a system is really cost effective if advancements of communica-

tion and computer engineering can be effectively utilized.

2 DESIGN CONSIDERATIONS

India is mainly a country with 70% population living in rural area with more than 40% population are living below poverty line and with 70% people are agriculturists. All the technical innovations in the field of home automation already popular in the developed countries cannot be directly suitable for the Indian conditions.

On basis of the above prerequisites we can set the following guidelines for the suitable home automation system for Indian homes. Indians cannot afford a costly home automation system. This discussion is focused on advanced idea of providing an integrated solution in which the comprehensive controlling and monitoring of all home appliances along with the security and smart energy management. The reason for integrating many things along with the home automation is one of the methods of cost reduction.

In India already mobile communication has become so popular that even the poorest families are using more than one mobile phone. So we find a solution to make the home automation a cost effective and easy to hand if we interface with already popular mobile communication system. The end user will be using a hand held device like mobile phone to control the home automation, home security and also do the energy management.

Home automation part mainly involves the things like light control, remote control, smart energy management, remote care, security and safety like burglar alarm, automatic water level control, automatic water temperature controller, automatic gate opening and closing, fire alarm, automatic emergency light controller, smoke detector, control of power generation using bio waste, automated electric fence for the farmlands, automatic operation of street lamp and gate lamp and night lamps, controller for intruder sensor.

Home security can be done with automatic biometric identification and authorization for entry, face recognition system, raising the alarm whenever unauthorized intrusion occurs, dialing to emergency helpline numbers during situations like

• Nagaraj Shet M.N., M.E., S.D.M Institute of Technology, Ujire-574 240, Mangalore (D.K), Karnataka, India, Mobile: +91 9241487640 E-mail: shet_nagaraj@yahoo.co.in.

• Shreesha Chokkadi., Ph.D., Professor & H.O.D., Dept of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal(Udupi dist)Karnataka Phone: +91 9448722353, E-mail: shreesha.c@manipal.edu

security breach, fire accidents, and untoward incidents. Energy management system involves management of the emergency lighting system, management of solar lighting system, management of micro wind power generation, power generation using hybrid power generation using biogas and solar etc. If every house can be made a power generating place using renewable energy sources and biomass energy generation systems, this will solve the energy shortage needs of the remote villages. India is already suffering a lot due to the transmission and distribution losses. Indian government instead of providing subsidized power to the farmers it can help them produce power generation from the biomass waste products. With this there can be way for income generation for the rural people.

The home automation and security and energy management system can integrate many other things like the health of the people, health monitoring of the old people, providing emergency medical help to the remote areas, monitoring of the government operated community health program related problems like pulse polio, malaria eradication, child vaccination, monitoring health and malnutrition. Health of every person in home from elderly to child can be monitored. This will reduce the wastage of government's efforts towards the social health related projects.

Government of India has many schemes to help the people who come under the category of below poverty line. Whether such benefits have reached those needy people or not can be effectively observed by having home automation system to monitor such activities also.

India has already started the ambitious project ADHAAR which provides the citizen ship card to all its citizens. This system can interface to the home automation system.

An integrated home automation system suitable to India requirements has multiple users and /or multiple stake holders. From the above discussion it can be concluded that many users are occupants of the house, family members, partners living in the same space, friends and family members living elsewhere who are involved in care or interested in well being of their family members, government representatives who are involved in the government schemes like social health workers, school teachers, doctors, people from government departments. Like this there can be many users. These people can be direct or indirect users or stake holders. Some of them can be in contact with data or people of home or can also be in direct need of the access to the devices of home care system.

Even though the users and stake holders are many there should be absolute privacy to the occupants of the house and there has to be serious measures taken to maintain the security of movable and immovable belongings of the members of the house.

Home automation system has to be flexible, easily modified to add additional facilities. This is required as the need of the people keep changing, the government policies keep changing, there can be additional facilities provided to improve the living conditions.

Next discussion will be related to the technical aspects. Tech-

nical aspects can be classified as things related to controller, remote devices, communication aspects, logic devices, sensors. Controller design aspects mainly can be open loop system or closed loop system. For all requirements one cannot make closed loop or open loop controller. This depends on the specific needs. We cannot involve more intelligence in the controller as this will increase the cost. It is better to have some set of preset parameters. The external inputs can be compared with the preset parameters and produce the signal to the actuators.

Sensors used in home automation can be wired or wireless sensors. If wireless sensors are used then lot of wiring will cause lot of cabling and long term maintenance will be problematic. As cost reduction measure all the sensors need not be wireless but some sensors can be made wired ones. But with wireless sensors we have to make the optimum analyses of the strength of the electromagnetic wave. Too much of electromagnetic interference will have to be avoided in the home environment. But an analysis can be made to make use of power line communication. So the existing electric wiring itself will be sufficient to connect the sensor devices present inside the home. So the method will be to connect all the sensors in the home through the wireline communication network.

There should be a centralized controller in the home which will communicate with the external network. This controller can be made to be enabled to communicate with the mobile type of hand held devices of the individual occupants and also it has the capability to interface with the local mobile tower or to the telephone line.

3 DESIGN AND IMPLEMENTATION

A basic in the rural atmosphere is an independent house with compound wall, this has been shown with the schematic as given in fig1. At specific places along the compound wall sensors are needed to be placed which can produce alarm during unauthorized intrusion. The gate has to be automated to provide access to authorized people only.

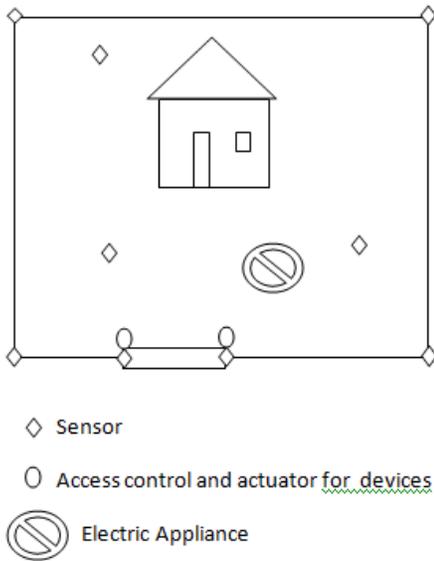


Fig 1. Layout plan of home automation setup.

As shown in the fig1 many sensors can also be placed at various places in the surrounding of the house. For example it may be for the purpose of knowing the humidity of the soil so that electric devices like the water pump can be switched on to start the sprinkler to water the plants or the lawn. At the gate of the compound sensors will give input to the controller and the actuator operating the gate can be driven by the output of the controller or by occupants of the house remotely by bypassing the controller also.

In fig.2 the typical house is shown. It is floor plan of a typical house. Every room will have a window opening to the outside. As shown sensors are placed at each window which is shown with sensor symbol. At all the doors sensors are placed along with the actuators to operate the doors. There are various electric devices in each of the rooms as shown with the symbol.

All the sensors and actuators can be connected to the central controller through a wired connection or wireless connection. The choice between these two implementations can be done depending on the cost and flexibility of implementation. More advanced implantation will be by making use of power line communication technique inside the house and surrounding to the house.

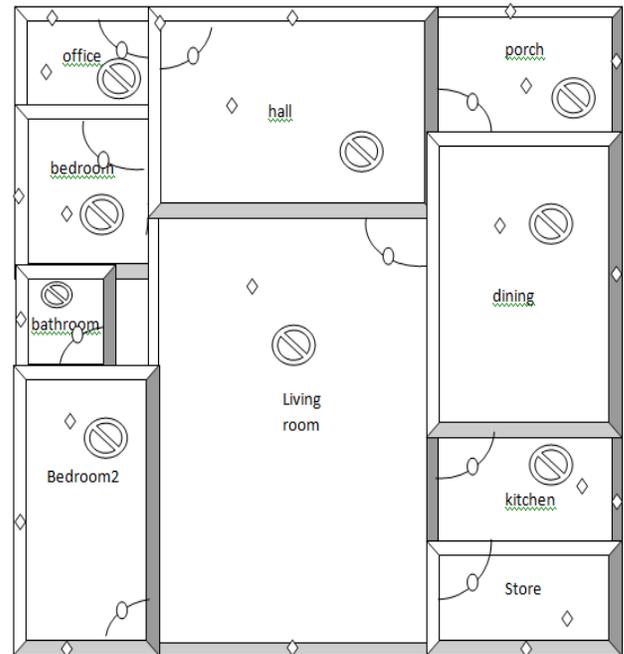


Fig2. Floor plan of home for home automation

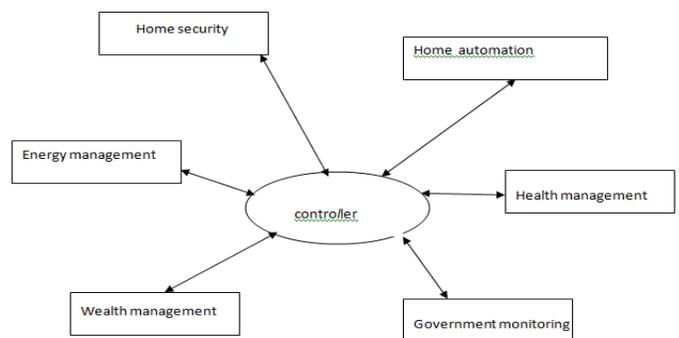


Fig3. Block diagram for integrated home automation

Fig 3 shows the implementation of the home automation suitable for the Indian home. Here the integration of the home automation home security, home energy management, health and wealth management and interfacing with the government agencies.

4 DISCUSSION

The above implementation gives the overall picture of integrated home automation system suitable for the Indian environment and the setup. There can be various types of implementations can be possible but the implementation should satisfy the needs of poor people, farmers, implementation of government policies, well being of elderly and also children.

5 CONCLUSION

This paper presents a new dimension to the concept of home automation. It will be not affordable for the Indian setup if only the home automation is done. But an integrated system as discussed in the paper will real beneficial if government comes forward and makes the coordinated effort by using the resources from energy department, telecommunication companies ,banking sector and also agencies involved in the social welfare and social justice. This is not a small task but a huge effort is required from lot many people.

REFERENCES

- [1] Carles Gomez and Josep Paradells, "Wireless Home Automation Networks: A Survey of Architectures and Technologies", IEEE Communications Magazine, June 2010, PP. 92-101.
- [2] Uzair Saeed, Sana Syed, Syed, Sheheryar Zahoor Qazi, Multi-advantage and security based Home Automation System, IEEE computer society, .UKSim 4th symposium on computer modeling and simulation.
- [3] Soyoung Hwang and Donghui Yu, Attribute-Based Routing for Home Automation Networks, IEEE International conference on Consumer Electronics 2011, PP.465-466
- [4] Mary Tome and Joaquin Sitte, Family System: A Reference Model for Developing Home Automation Applications, 2006 IEEE International Conference on systems, Man and Cybernetics October 8-11, 2006, PP 32-37
- [5] Folco Fioretti, Paola Pierleoni, Samuele Pasqualini, On use of xAP Home Automation Protocol for networking in critical environments, Feb 17-20, 2008, ICAT 2008, PP1960-1965.
- [6] Taewan Kim, Hakjoon Lee and Yunmo Chung, "Advanced Universal Remote Controller for Home Automation and Security", IEEE 2010, PP.2537-2542
- [7] Manouchehr Ghahramanian Golzar, Hamid Reza Tajozakerin, "A New Intelligent Remote Control System for Home Automation and Reduce Energy Consumption" IEEE computer Society, 4th International Conference on Mathematical/ Analysis Modelling and Computer Simulation, 2010, PP.174-181

Comparison between EOR methods (Gas Injection, Water injection and WAG Processes) in One of Iranian Fractured Oil Reservoirs

Maliheh Alsadat Mousavifar¹, Riyaz Kharrat², Ashkan Parchizadeh³, Sedigheh Mahdavi⁴

Abstract: In the common enhanced oil recovery methods, unsuitable mobility ratio within the injected fluid and the oil in the reservoir, leads to reduction of sweep efficiency especially in fractured reservoirs. In this work, gas injection, water flooding and WAG processes were applied in well configuration which called Dual-Five spot. Because this sector model was selected from a highly fracture reservoir. Dual-Five spot with 38.889% oil recoveries are the best EOR plans for this sector model. The main important point in using WAG process in this sector model is that the cycle of injected gas should be less than injected water.

Keywords: WAG, Well pattern, Simulation, Sweep efficiency, Optimizing, WAG Cycle, WAG Ratio

1. Introduction

For enhanced oil recovery purpose, miscible gas flooding and water alternating gas (WAG) process have been applied successfully in many hydrocarbon reservoirs [McGuire P.L., et al., 2003, Thomas F.B., 1994]. The attractiveness of miscibility is that it can reduce the interfacial tension. The reduction in the interfacial tension has a significant effect on relative permeabilities and residual saturation by increasing the trapping number, which has been formulated mathematically and tested by experiments [Qiliang, B., et al., 2003, Wagner, O.R., et al., 1996].

The WAG process is designed to improve the continuous gas injection EOR method, mainly by reducing gas mobility and thereby increasing sweep efficiency in the reservoir. For any hydrocarbon reservoir, all causes of inefficient oil recovery can be formally organized into just five factors:

1. Heterogeneity and gravity effects. 2. Fractional flow effects. 3. Local or displacement sweep efficiency. 4. Capillary trapping of residual oil--residual oil saturation. 5. Streamline effects.

To maximize the oil recovery, we need to enhance displacement sweep efficiency and the volumetric sweep efficiency, which is defined as the product of areal sweep efficiency E_A and vertical sweep efficiency E_I , or to reduce the residual oil saturation [X. WU., et al., 2004]. The reservoir sweep efficiency to be expected from a gas injection is to a large extent dominated by the degree of heterogeneity if the displacement is not fully miscible. Analysis of historical water injection performance data will typically indicate presence of high-perm streaks and fracture network, which would prove harmful

to the microscopic sweep efficiency in a gas injection scenario by causing gas channeling and premature breakthrough [K. Mogensen, et al., 2010]. It is fundamental to the study of different displacement processes in reservoir systems knowing how the multiphase mixtures flow in response to heterogeneity. Small-scale heterogeneities are particularly problematic for all secondary and tertiary recovery processes as they can cause distortion of fluid streamlines and deviation from the production profiles of equivalent homogenous systems [Y.M. Al-Wahaibi, et al., 2011]

This paper is one of the few quantitative studies of the gas injection, water injection and WAG process including evaluation of parameters affecting project design, selection of injection and production well location and optimization of production well controls at gas injection, water flooding and WAG processes.

2. Fluid Properties

This field is located in southwestern of Iran. Water cut and gas oil ratio was set 90% and 700 ft³/scf respectively. The other fluid properties are listed in Table 1. Figure (1) illustrates the final matching of phase diagram of reservoir fluid. Reservoir temperature is 115°F.

¹ Corresponding Author, E-Mail: msm.ir2010@yahoo.com, Mailing Address: Department of Oil and Gas Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

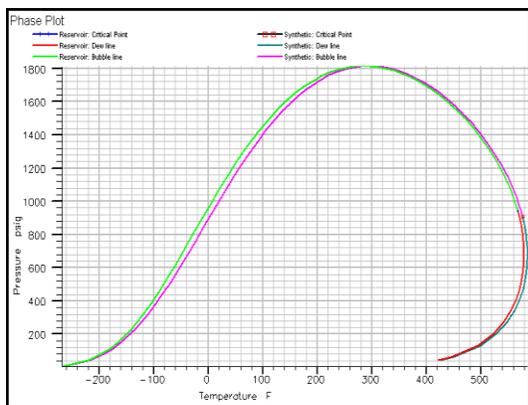


Figure (1): Phase diagram of reservoir fluid

3. Reservoir Rock Properties

Initially the oil reservoir was under saturated and initial reservoir pressure was equal to 2500 psia, Oil production causes to reduction in pressure. After reaching to saturation pressure gas cap started to form gradually. Reservoir formations con-

sist of grey and brown dolomitic and carbonate dolomitic rocks with oil field fissures and fractures. This zone can be divided into two subzones. Upper subzone which is carbonate has an average thickness of 53 meters. Lower subzone consists mostly of alternating carbonate and anhydrite rocks with average thickness of 16 meters. Fracture study implicated network is fairly distributed which includes most of the oil in place and these fracture networks are contributing to oil production which shows the reservoir is acting as dual porosity continuum.

4. Sector Model Properties

In this paper with regard to design the suitable well configuration and EOR method, the different injection scenarios with different well patterns were designed using commercial simulator. The properties of selected sector are listed in Table 1.

Table 1: Fluid and sector model properties

Fluid properties	API	39	Oil FVF, RBBL/STB	1.34
	GOR, ft3/scf	700	Water FVF,	1.01
	Water compressibility, 1/psi	2.12E-06	Oil viscosity, cp	0.65
	Oil density, lbm/ft3	45	Gas viscosity, cp	0.019
	Gas density, lbm/ft3	0.049	Water viscosity, cp	0.18
Sector model properties	Type of porous medium	Fractured	X grid block size, ft	2180
	Number of cell in X-direction (N_x)	33	Y grid block size, ft	1130
	Number of cell in Y-direction (N_y)	31	Z grid block size, ft	116
	Number of cell in Z-direction (N_z)	7	Matrix porosity, %	7
	Number of cell	7161	Fracture permeability, md	5800
	Dual porosity matrix-fracture coupling, 1/ft ²	0.6	Effective matrix block height for gravity drainage, ft	20

Based on the Geological data, fracture media exists, so dual porosity model was chosen for simulation of reservoir. Then reservoir static data including porosity, absolute permeability and NTG were calculated for all grid blocks using geological model of reservoir and up scaling techniques and were used as input data to simulator. In order to classify reservoir rock, first using frequency diagram of initial classification of various kinds of rock-water saturation diagram, porosity and the results of special experiments and various kinds of rock were divided into four groups. Then by drawing water saturation diagram, porosity and applying water saturation and porosity sections, reservoir rock was divided into six types of rock. Figure 2(a) shows the selected sector.

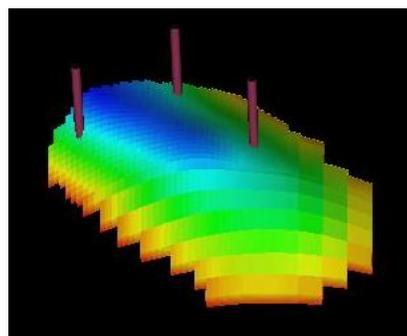


Figure 2 (a): Three dimension view of selected sector with three production wells

5. Results

5.1. Natural Depletion

Selected sector model has three production wells; rate of field oil production was set at 6000 STB/DAY with divided between three wells. This natural depletion was called (Nd-DONO-6000). According to the result of do noting scenario, ultimate oil efficiency was obtained only 5/9%. After adding three other production wells with the same field oil production rate, efficiency reached to 27.24 %. Pay much attention to the location of infill drilled wells is necessary.

With regard to obtain the best location of injection wells, Dual-Five spot patterns have six production wells and two injection wells (figure 2(b)). Because of strong and dominate fracture network, the location of wells especially injection wells has a huge effect on recovering of fluid in place.

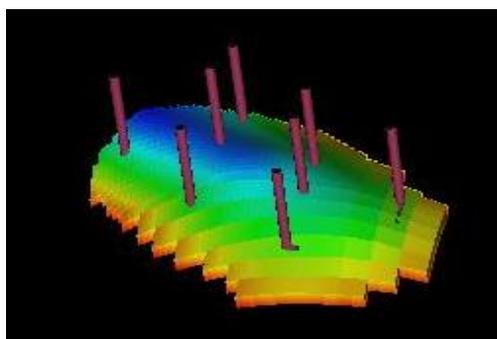


Figure 2(b): Three dimension view of selecte sector with six production wells & two injection wells

5.3. Gas Injection

In this part, several gas injection scenarios with different injected gas rate during 1.5 to 60 MMSCF were applied to reach the optimum gas rate. With increasing the rate of gas injection, field oil efficiency increased but there is no clear difference between higher rates. Figure 3 illustrates maximum oil recovery was obtained 32.003%.

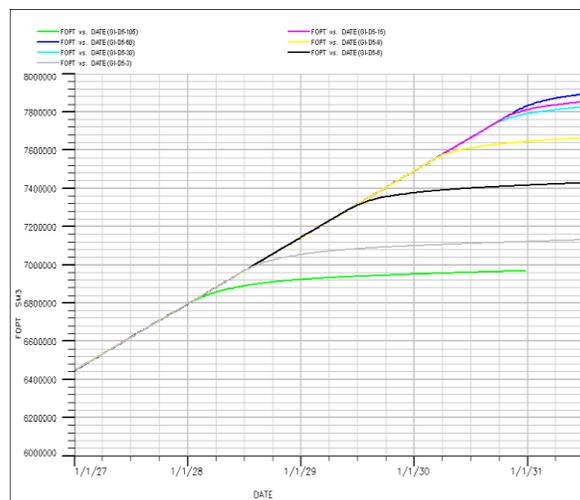


Figure 3: Field oil production total at gas injection scenario at Dual-Five spot (D5)

5.4. Water Flooding

Figure 4 shows the effect of water injection rate (5000 to 30000 BBL/DAY) on field oil production of selected sector model. As it is seen scenario of water injection with 30000(bbl/day) has higher efficiency in this pattern. Beacouse of highly dencity of fracture in this sector, efficiency in water injection scenario is better than Gas injectuin scenario.

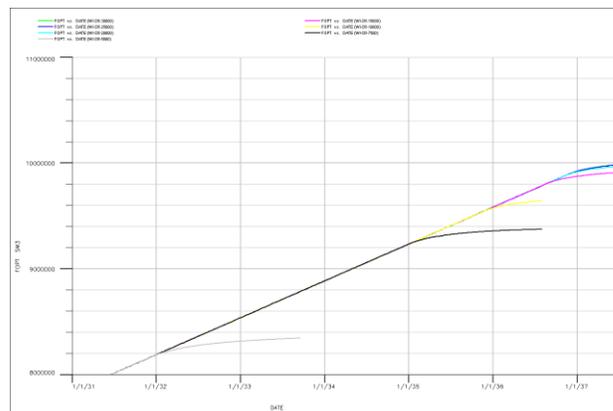


Figure 4: Field Oil Production at water injection scenario at Dual-Five spot pattern

5.5. WAG Flooding

In WAG process there are some important parameters which included WAG ratio and WAG cycle that planned WAG ratios are 1:1, 2:1, 3:1, 4:1, 1:2, 1:3. the first number of ratio correspond to water and the second number to gas. First two pore volume injections 0.1 PVI and 0.5 PVI were considered. The results of different WAG pore volumes

injections and WAG ratios are listed in Table 2.

Table 2: Effect of pore volume injection and WAG ratios on Dual-Five spot pattern at WAG process

Name	Qw(bb/day)	Qg(MMSCFD)	FOE%	FOPT(SM3)
WAG-0.1PV-31	22500	20	38.713	9556371
WAG-0.1PV-21	15000	20	38.581	9523566
WAG-0.1PV-13	7500	60	36.483	9017749
WAG-0.1PV-12	7500	40	36.539	9025493
WAG-0.1PV-11	7500	20	36.941	9134618
WAG-0.1PV-41	30000	20	38.733	9563569
WAG-0.2PV-41	40000	30	38.729	9560369
WAG-0.2PV-31	30000	30	38.608	9530676
WAG-0.2PV-22	20000	30	38.525	9522115
WAG-0.2PV-13	10000	90	37.05	9145539
WAG-0.2PV-12	10000	60	37.372	9213045
WAG-0.2PV-11	10000	30	37.628	9290653

First different pore volume injections and WAG ratios were optimized in designed patterns, in next step effect of WAG cycle on WAG process was investigated. Several WAG cycles (6, 10 and 12 month) were applied, each of WAG cycle time can divided to two parts which the first one is the amount of gas injection time and the other one corresponds to water injection time. According to Figure 5, the highest oil recovery at Dual-Five spot pattern, WAG-D5-0.1PV-41-210 (38.889%), also effect of WAG cycle at Dual-Five spot pattern with different ratios in Figure 5 was displayed. Comparison between all of scenarios at 0.1 and 0.2 pore volume injections at different cycles explains that 0.1 PVI is more suitable to reach higher oil recovery; the reason of this occurrence is function of network fracture. The greatest efficiencies, 38.889 % were achieved at WAG-D5-0.1PV-41-210 and 38.73 % at WAG-D5-0.2PV-41-210.

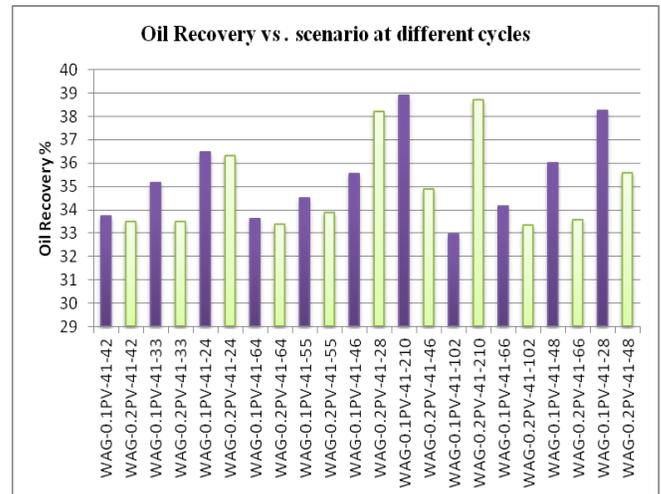


Figure 5: Comparison between WAG cycle with 0.2 and 0.1 PVI at Dual-Five spot pattern

With consider to Figures 5, all of implemented scenarios at 0.1 PVI have higher efficiency than 0.2 PVI totally. Figure 6 describes the total result of all of the best scenarios. The first part is included of gas injection scenarios, so selecting gas injection method lead to 32 % efficiency.

For applying WAG process, Dual-Five spot scenarios with 38.889 % oil recovery can select as the best method, But sensitivity analysis on location of injection and production wells shows the location of wells is more important in WAG rather than gas injection. At all of well configurations, water injection has the first rank. There are two injected wells, it is necessary to know that the cumulative of water injection is constant. Number of injection well clearly affects on oil recovery because cause to spread fluid in a stable movement especially in-fractured media.

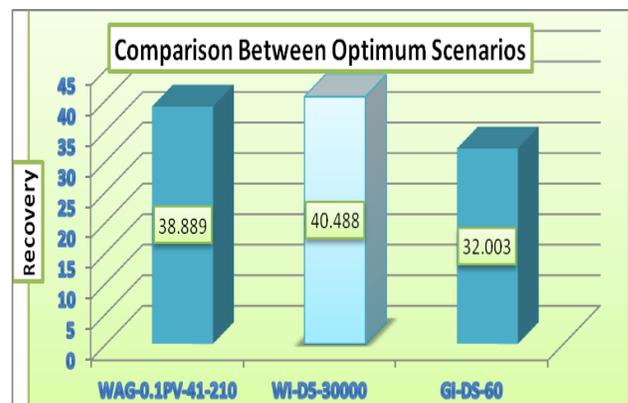


Figure 6: Comparison of oil recovery between Optimum EOR method (gas injection, water injection, WAG)

6. Conclusions

Base on the results, intensive network fracture has skirted the model and fracture veins have strong relation, therefore gas can move rapidly and easily inside the reservoir and lead to minor oil recovery. The main important point in using WAG process in this sector model is to adjust period of gas injection time. Whatever the cycle of injected gas is less than injected water, model will visit the limitation of GOR subsequent. The other significant part in WAG process is number of cycles. If the number of cycle would be less it tends to steady and suitable sweep efficiency for this reservoir.

References

- [1] McGuire P.L. Holt B.M, Unconventional Miscible EOR experience at Prudhoe Bay: A Project Summary; SPE 82140, SPE Reservoir Evaluation & Engineering, Feb. 2003
- [2] Thomas, F.B., Holowach, N., Zhou, X.L., Bennion, D.B., Bennion D.W., Miscible or Near Miscible gas Injection, which is better; SPE/DOE 27811, Improved Oil Recovery, Tulsa, Oklahoma, U.S. A., April 1994.
- [3] Qiliang, B., Zhong, Taixian, Z., T., Qingjie, L, Pore Scale Network Modeling of Relative Permeability in Chemical Flooding; SPE 84906, International Improved Oil Recovery Conference, Kuala Lumpur, Malaysia, October, 2003
- [4] Wagner, O.R, Leach, Ro.O. Effect of Interfacial tension on displacement efficiency; Soc. Pet. Eng. J., 6(Sep.):335-344, 1996
- [5] WU, X., Ogbé, D.O., ZHU, T., Kataniar, S., Critical Design Factors and Evaluation of Recovery Performance of Miscible Displacement and WAG Process. Canadian International Petroleum (CIPC), June 2004.
- [6] K. Mogensen, P. Hood, and Robert Jones, R. Noman, Gas Injection Project in the KharaiB B Reservoir of the Giant Al Shaheen Field, SPE 129565, SPE EOR Conference at Oil & Gas West Asia, Muscat, Oman, April 2010.
- [7] Y.M. Al-Wahaibi, A.K. Al-Hadrami, Lenses Heterogeneity Effects on Water Alternating Gas Injection in Oil Wet Porous Media, SPE Projects and Facilities Challenges Conference at METS Doha, Qatar, February 2011.

Implementation of Data Encryption & Decryption For the Safer+ Algorithm Using Verilog HDL

J.Umesh rao, Dr.E.Nagabhooshanam

Abstract— A VLSI implementation for the SAFER+ encryption algorithm is presented. The combination of security and high speed implementation makes SAFER+ a very good choice for wireless systems. The SAFER+ algorithm is a basic component in the authentication Bluetooth mechanism. The relation between the algorithm properties and the VLSI architecture are described. The whole design was captured entirely in VHDL using a bottom-up design and verification methodology. A FPGA device was used for the hardware implementation of the algorithm. The proposed VLSI implementation of the SAFER+ algorithm reduces the covered area about 25 percent, and achieves a data throughput up to 320 Mbps at a clock frequency of 20 MHz and proposed architecture high data throughput of 704Mbits/sec at a maximum clock frequency of 44MHz, at a cost of area reduced.

Index Terms— Bluetooth Mechanism, Cryptographic, Decryption, Encryption, Proposed Architecture, Safer+, Simulation results.

1 INTRODUCTION

Wireless communication technology has advanced at a very fast pace during the last years, creating new applications and opportunities. In addition, the number of computing and telecommunications devices is increasing. Special attention has to be given in order to connect efficiently these devices. In the past, cable and infrared light connectivity methods were used. The cable solution is complicated since it requires special connectors, cables and space. This produces a lot of malfunctions and connectivity problems. The infrared solution requires line of sight. In order to solve these problems a new technology, named Bluetooth [1], [2], has been developed. With this communication system, users are able to connect a wide range of computing and telecommunications devices easily and simply, without the need for connecting cables. Bluetooth is a technology and standard, designed as a wireless-cable replacement to connect a wide range of devices. Unlike wireless LANs such as 802.11b, it was designed to be low power, operate over a short range, and support both data and voice services. It enables peer-to-peer communications among many types of handheld and mobile devices. Furthermore, it provides a conceptually simple communication model and lets these devices exchange information and work together to benefit the user. The aim of this project is to develop a safer+ algorithm which achieves a high data throughput. The approach taken will be a prototype mechanism in Verilog and simulate the same.

2 SAFER +

- JARUPULA UMESH RAO received the B.Tech degree in ECE from VJIT, M.Tech in VLSI system design from Aurora scientific and technical research academy, Hyderabad. He is currently pursuing research in VLSI design from rayalaseema university, Kurnool. He has 7 years teaching experience, presently he is working as Associate professor in Nigama engineering college in the department of ECE. He is a life member of ISTE, he has published and presented 6 research papers in various reputed international/national journals and conferences respectively. PH: +91-9676373876, E-mail: umesh481@gmail.com

The SAFER+ (Secure and Fast Encryption Routine) algorithm is based on the existing SAFER family of ciphers, which comprises the ciphers SAFER K-64, SAFER K-128, SAFER SK-128 bits data [2]. They have been developed by James L. Massey at the ETH Zurich [4], [5], [6]. SAFER+ (as is also the case with all prior ciphers in the SAFER family) is neither a Feistel cipher nor a substitution-permutation cipher. There is no fundamental reason to alternate between substitutions and permutations to create good confusion and diffusion. All algorithms are byte-oriented block encryption algorithms, which are characterized by the following two properties. First, they use a non-orthodox linear transformation, which is called Pseudo-Handmaid-Transformation (PHT) for the desired diffusion, and second, they use additive constant factors (Bias vectors) in the scheduling for weak keys avoidance.

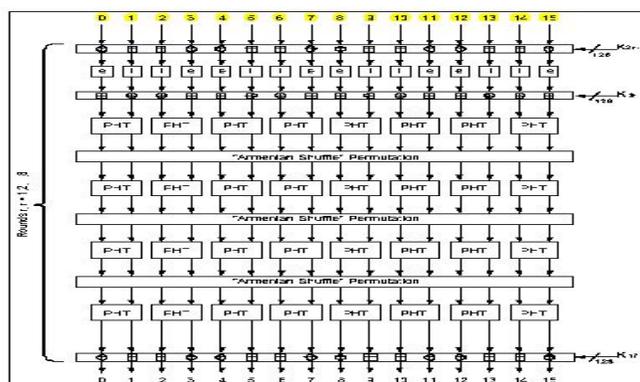


Fig. 1. One SAFER+ Encryption Round

2.1 Cryptographic Strength of Safer+

Differential cryptanalysis [2] has proved to be the most effective

tive general attack the previous SAFER family of ciphers, and appears also to be the most effective general attack against SAFER+. This task of showing the security of r-round SAFER+ against differential cryptanalysis is essentially that there are no (r-1)-round characteristics with probability greater than 2^{-128} . An exhaustive study of SAFER+ has shown that all 5-round characteristics have probability significantly smaller than 2^{-128} . The conclusion is that SAFER+ with six or more rounds is secure against differential cryptanalysis. SAFER+ enjoys good diffusion (i.e. to ensure that small changes in round inputs cause large changes in round outputs) also ensures that "differences" similarly propagate and is the main source of the strength of SAFER+ against differential cryptanalysis.

3 ARCHITECTURE OF SAFER+ ALGORITHM

The architecture for the implementation of the SAFER+ algorithm [2] consists of the two main components, the data encryption path and the key scheduling. The plain text passes through the r rounds of encryption where r is determined by the key length chosen for the encryption. In our implementation we are using key size is 128 bits (fig.1), so the no of rounds becomes eight. Two 16-byte round sub keys are used within the each round of encryption. These round sub keys are determined from the user-selected key according to a key scheduling. Finally the last round sub key "2r+1" is to Mixed Xor/Byte -Addition with the r rounds of encryption. This addition constitutes the output transformation for safer+ encryption. The input for the decryption of the safer+ is the cipher text block of 16-bytes. The decryption begins with the input transformation that undoes the output transform in the encryption process. This block then process through the r rounds of decryption, round1 of which undoes the r round of encryption, round r undoes the encryption of round1 of encryption to produce the original plaintext. The round sub keys used for decryption used same as encryption but applied in reverse order.

The decrypting structure [4] of SAFER+ is shown in Fig.3. The deciphering algorithm consists of an input transformation that is applied to the cipher text block, followed by r rounds of identical Transformations. The input transformation consists of the Mixed XOR/Byte-Subtraction of sub key K_{2r+1} from the cipher text block. A characterizing feature of SAFER+ is that decrypting rounds differ from encrypting rounds so that an encrypted cannot be converted to a decrypted by simply reversing the key schedule. The output of input transformation which undergoes the 8-rounds of decryption. The first round of decryption undoes the r round of encryption. The keys are used same as encryption but applied in reverse order.

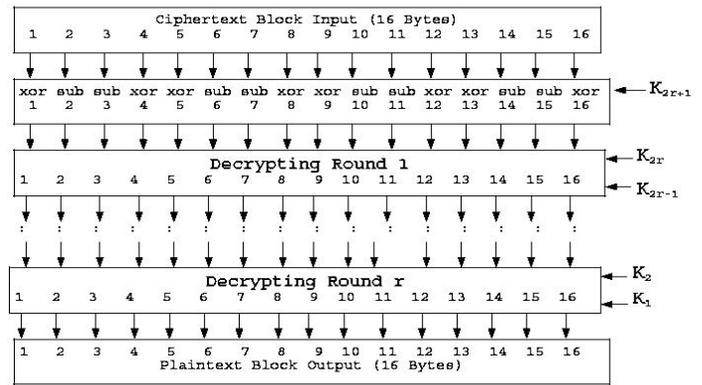


Fig. 3. The SAFER+ Decryption

4 MODIFIED ARCHITECTURE

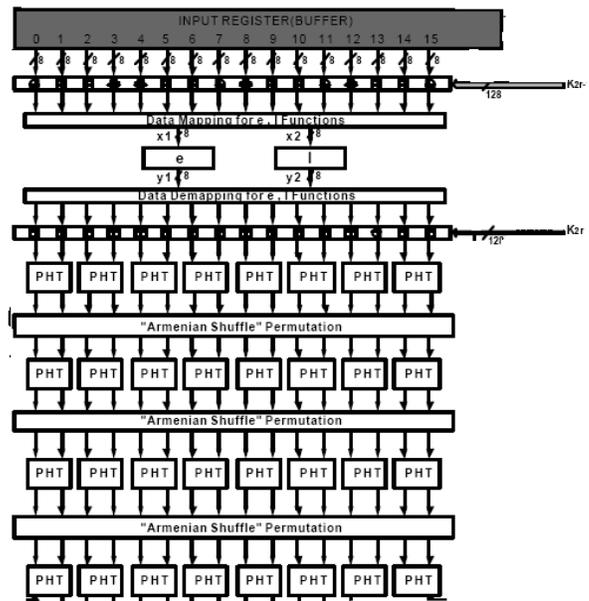


Fig. 4. Modified Architecture

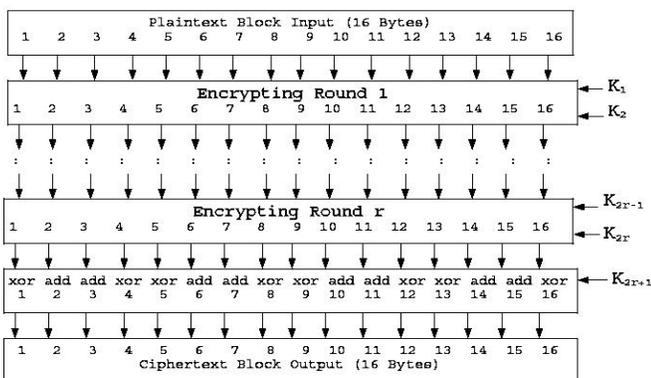


Fig. 2. The SAFER+ Encryption

3.1 Decryption of Safer+

The modified single round implementation is chosen because the required system throughput can be achieved and in the

same time the covered area is minimized. In this modified architecture we use the concept of data mapping and damping. The damping unit performs the reverse function of the data mapping. This design results in reduction of the covered area than conventional implementations. After that the output of the nonlinear layer is added by a mixed byte-addition/xor with a round key.

The operations after that are four Pseudo-Handmaid-Transform (PHT) layer. That is connected through by three permutations. The decryption operation is reverse to the encryption operation. The encryption and decryption structure is dissimilar. In the decryption process the keys are applied in reverse order compared to the encryption process.

5 SAFER+ ENCRYPTION

In this implementation entire design has been divided in to various modules given below:

1. Safer encryption
2. Safer single
3. mod_add
4. xor_bit
5. e_block
6. l_block
7. permutation
8. pht

5.1 SAFER+ Encryption Implementation

SAFER+ algorithm encryption [6]. Implementation has been implemented as top level module. All other modules (safer single, modular addition, Bit wise ex-or, 'e' and 'l' blocks, permutation boxes, and Pseudo Handmaid Transform (PHT)) have been called in this top level module. The main block takes 128-bit key and 128-bit plain text as inputs and output will be 128-bit cipher.

5.2 SAFER+ single round implementation

In this proposed design the whole single round of the SAFER+ algorithm is implemented. In order to run the whole SAFER+ algorithm eight loops of the single round implementation are needed. The single round implementation [6] is chosen because the required system throughput can be achieved and in the same time the covered area is minimized. This block takes two 128 bit keys and 128-bit plain text as inputs and output will be 128-cipher.

5.3 Modular addition

Safer+ algorithm involves four layers of 8-bit modular additions. Modular adders and bitwise ex-or are interleaved alternatively in each of the four layers. This modular addition is performed over GF (256).

5.4 Bit EX-OR

Bit-wise ex-or blocks are also used in the single round of safer+ algorithm in combination with modular addition blocks.

5.5 'E' and 'L' Blocks

Substitution box layer introduces non-linearity to the safer+ algorithm which is an essential feature in any of the security algorithms. Substitution box contains 'e' and 'l' non-linear functions and have been defined as follows:

$$e, l : \{0, \dots, 255\} \rightarrow \{0, \dots, 255\},$$

$$e : i \rightarrow (45i \pmod{257}) \pmod{256},$$

$$l : i \rightarrow j \text{ such that } i = e(j).$$

In total eight 'e' and 'l' blocks are required for the algorithm. In the hardware implementation, to minimize the area only one set of 'e' and 'l' blocks are used

5.6 PHT Round

The four linear PHT layers connected through the permutations as shown in Figure.5

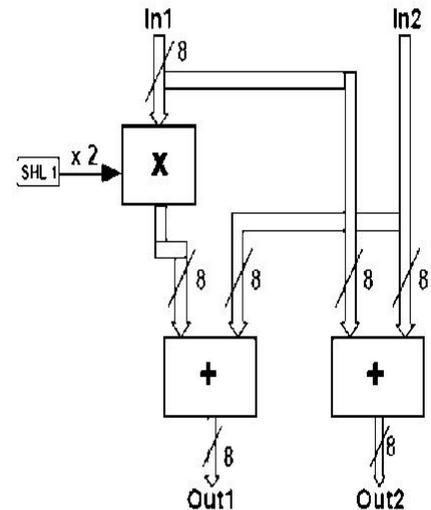


Fig. 5. Permutation boxes

PHT stands for Pseudo Handmaid Transform. The PHT boxes defined as

$$PHT (in1, in2) = (2in1 + in2, in1 + in2).$$

The outputs of the PHT,
 out1 = 2in1 + in2

$$out2 = in1 + in2 \text{ are implemented in GF (256).}$$

6 SAFER+ DECRYPTION

In this implementation entire design has been divided in to

various modules given below.

1. Safer decryption
2. Safer_desingle
3. Mod_subtract
4. Inverse permutation
5. Inverse pht

6.1 SAFER+ Decryption Implementation

Safer+ algorithm decryption implementation has been implemented as top level module. All other modules (safer+_desingle) modular subtraction, Bit wise ex-or, 'e' and 'l' blocks, inverse permutation boxes, and inverse Pseudo Handmaid Transform (IPHT)) have been called in this top level module. The main block takes 128-bit key and 128-bit plain text as inputs and output will be 128-bit cipher.

6.2 IPHT Block

IPHT stands for Inverse Pseudo Handmaid Transform. The IPHT boxes defined as the outputs of the IPHT,

$$\text{out1} = \text{in1} - \text{in2}$$

$$\text{out2} = -\text{in1} + 2\text{in2} \text{ are implemented in GF (256).}$$

Single IPHT block implementation is shown in Figure 3.9. In IPHT block Multiplication by 2 can be achieved by one bit left wired shift.

In the each single round of an encryption consists of a four pht blocks and three blocks of permutations. Permutation is after the each pht block. The permutation block performs the change the byte positions which are came from pht block.

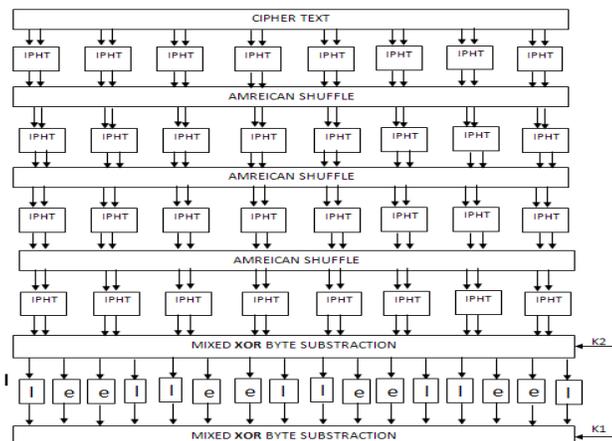


Fig. 6. The IPHT Implementation

The permutation box performs the how the input bytes are indices are mapped into output bytes. Thus Position 0(left most) is mapped on 8 byte; byte 1 is mapped on 11 byte like

that the permutation box performs the operation. In the process of decryption the permutation box performs the reverse the operation of an encryption permutation. Due to this reverse permutation in decryption causes the same positions in actual plaintext.

7 VERILOG HDL

Implementation of High Speed CRC is done using Verilog HDL. In the semiconductor and electronic design industry, Verilog is a hardware description language (HDL) used to model electronic systems. Verilog HDL, not to be confused with VHDL, is most commonly used in the design, verification, and implementation of digital logic chips at the Register transfer level (RTL) level of abstraction. It is also used in the verification of analog and mixed-signal circuits.

7.1 Experimental and Simulation Results

The whole design was captured entirely in verilog HDL language. All of the system components have been described with structural architecture. The proposed architecture is synthesized by using FPGA device of XILINX [7].

Final Timing Optimization Statistics for the design

- CLK : 44 MHz
- FPGA used for synthesis : VIRTEX IV-PRO
- Devices used : 12ff152
- Throughput : 704Mbits/sec

TABLE 1

COMPARISON BETWEEN PREVIOUS AND MODIFIED ARCHITECTURE

Type	Previous	Proposed
Gate level count required	233839	200013
Frequency	20 MHz	44 MHz
Throughput	320Mbits/sec	704Mbits/sec

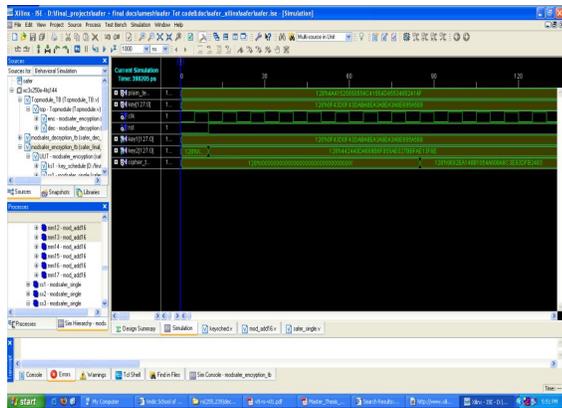


Fig. 7. Simulation results of Safer+ encryption

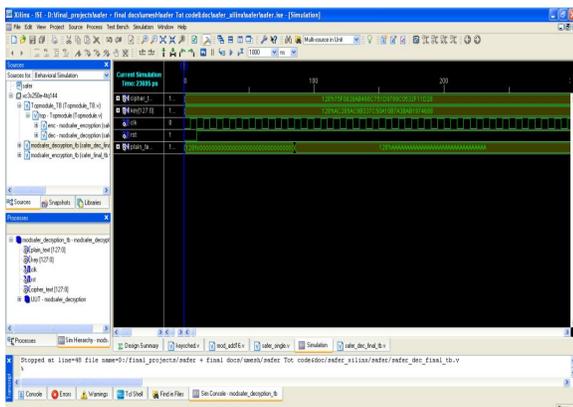


Fig. 8. Simulation results of Safer+ decryption

8 CONCLUSION

In this project, implementation of Safer+ algorithm (which is most important algorithm) has been carried out successfully has been done. This project has helped me to become familiar with Verilog HDL, simulation tools (Incisive (TM) unified simulator©5.6 and Modelsim© 6.0E) and various synthesis tools (Encounter RTL Compiler-XL©Cadence Mentor Graphics© FPGA Advantage and Xilinx Web pack ISE 6.1i). FPGA device has been used for the implementation of the algorithm. VLSI implementation of the SAFER+ algorithm has been observed to work with a high throughput of 704Mbits/sec at a maximum clock frequency of 44MHz, at a cost of area reduced. Measurement results and comparisons between the proposed and previous implementations are presented.

ACKNOWLEDGEMENT

The author would like to acknowledge the support of the author's supervisor, professor and head Dr.E. Nagabhooshnam (MGIT) and Mr.S.Ramesh babu (NEC) for their continuous interest, encouragement and guidance through out the research. This research is supported by Rayalaseema univer sity, Kurnool and Nigama engineering college, karim ngar.The author would also like to thank everybody who was involved directly or indirectly in the research project, especially the author would like to express his heartfelt thanks for the support given by the secrecy and joint secrecy of NEC, karimnagar.

REFERENCES

- [1] "Specification of the Bluetooth System", Specification Volume1, Version 1.1, February 22, 2001.
- [2] J.L. Massey, G. H. Khachaturian, M. K. Kuregian, "Nomination of SAFER+ as Candidate Algorithm for the Advance Encryption Standard", *First Advanced Encryption Standard Candidate Conference*, Ventura, CA, August 20-22, 1998.
- [3] J. L. Massey, "On the Optimality of SAFER+ Diffusion", *Second Advanced Encryption Standard Candidate Conference (AES2)*, Rome, Italy, March 22-23, on line available at <http://csrc.nist.gov/encryption/aes/round1/conf2/aes2conf.htm>.
- [4] J. L. Massey, "SAFER K-64: A Byte-Oriented Block Ciphering Algorithm", *Fast Software Encryption*, Proceedings of the Cambridge Security Workshop, Cambridge, U.K, 1998, pp. 1-17.
- [5] P. Kitsos, N. Sklavos and O. Koufopavlou" *HARDWARE IMPLEMENTATION OF THE SAFER ENCRYPTION ALGORITHM FOR THE BLUETOOTH SYSTEM*" Vol. IV, pp. 878-881, USA, May 26-29, 2002
- [6] A. Schubert, V. Meyer, W. Anheier, "Reusable Cryptographic VLSI Core Based on the SAFER K-128 Algorithm with 251,8 Mbits/s Throughput", *IEEE Workshop on Signal Processing Systems*,1998,pp. 437-446
- [7] Xilinx, San Jose, California, USA, Vertex, 2.5 V Field Programmable Gate Arrays, 2001, www.xilinx.com

The Impact of Gas Flaring and Power Plant Emission on the Socio Economic Environment of Niger Delta Region in Nigeria

Atoyebi, Kehinde .O. & Akinde, Jubril Olatunbosun

ABSTRACT: Nigeria flares 17.3 billion m³ annually from the crude oil exploration in the Niger Delta. This gas flaring expends huge amount of energy and causes environmental degradation and disease. The links between oil exploration and exploitation processes and attendant environmental, health and social problems in oil producing communities are not well known.

This study examines how gas flaring during oil exploration has caused hardship to the local communities in the Niger Delta region of Nigeria. Nigeria which is known to be the sixth largest producer of crude oil among the organization of petroleum exporting countries (OPEC) and the largest in the African continent.

Despite this enormous wealth coming from the Niger Delta, There is a pervasive poverty and despicable environmental damage as a result of crude oil mining activities going on in the region. The neglect of the oil communities by the government and the oil companies in terms of infrastructural development and youth empowerment has led to bunkering, restiveness and hostage taking in the area. This study made of primary and secondary data to analyse the discussion contained in the study. It identifies the constraints to effective implementations of Nigeria's environmental laws and especially the Environmental Impact Assessment (EIA) as it concerns oil prospecting which hindered environmental sustainability. It also Unveils critical issues concerning the region deplorable state of the economy of the local people of the Niger Delta in Nigeria, which has the stigma of being the world's highest gas flare of the gas associated with crude oil production. Apart from being wastage of valuable resources, it runs contrary to Nigeria's obligations to reduce green house gas (GHG) emission.

This paper also reviews the trend of gas flaring over the years (ii) reviews the Nigerians national response and its commitment to end flares by 2008 and the prospects of the Nigeria's Liquefied Natural Gas (LNG) projects.

The results show that the reduction of gas flaring will improve human health and the environment. This paper concludes that the livelihood of the people in the Niger Delta can be improved by promoting a shift from flaring the associated gas to utilization as a gaseous fuel for electricity generation.

Keywords: Gas flaring, degradation, sustainability, environment.

1.0

INTRODUCTION

Nigeria is richly endowed with energy resources; these include coal, tar sand, oil, natural gas, hydroelectricity, solar and so on. The commercial energy sector is, however, dominated by oil and gas, both of which jointly account for 71 per cent of commercial domestic energy resources (Iwayemi and Adenikinju, 2001). Thus, oil and gas play significant role in the development of the Nigerian Economy.

The discovery and extraction of natural resources has brought different consequences to countries that are endowed with such resources. While some of these nations have become economically strong and self sustaining, others have been drawn into serious economic hardships and conflicts. Empirical evidence has revealed that oil and gas abundant economies are among the least growing economies (Ayres & Kneeece (1990), Barnes 2005; DEckor, 2002. This Phenomenon of often conceived within the prisms of the "resource curse" and "Dutch disease" but both are manifestatives of the inefficient utilization of resources rather than inevitable out come of the availability of oil gas resources. United Nations Development Programme (UNDP, 2006) the proponents of the resource curse, project have it that the citizens of these countries rather suffer from abject poverty, environmental damages, pollutions, diseases, illiteracy and score very low on the United Nations Human Development Index.

It is established that economic advancement and industrialization are contingent on continuous availability and prudent utilization of energy resources, still Nigeria's energy planning and management system leaves much to be desired. Nigeria despite its widely acknowledged status as an energy rich nation still saddled with the problem of adequate and unreliable energy supply for domestic and industrial use (NCEMA, 1999).

The Niger Delta region, where Nigeria Current Large Oil and Gas resources are located, to with the Niger Delta as the unifying feature has remained a source of global interest. With openness to the Atlantic Ocean and watercourses with access to the sea and rivers such as the Benue and Niger Rivers, the Niger Delta embodies some of the major coastal upwelling sub-ecosystems of the world and is an important center of marine biodiversity and marine food production ranked among the most productive coastal and offshore waters in the world. However, pollution from domestic and industry sources, over-exploitation of Oil and Gas resources and poorly planned and managed communities and coastal developments and near-shore activities are resulting in a rapid degradation of vulnerable land, coastal and offshore habitats and shared living marine resources of the region putting the economies and health of the populace at risk.

The deterioration in water and air quality (chronic and catastrophic) from land and sea-based activities (especially industrial, (flaring/power plants), agricultural, urban and domestic sewage run-off, eutrophication and gas flaring have been identified as a major Transboundary environmental problem by communities in the region.

An in-depth study of the Natural gas in Nigeria becomes imperative because of its pervasive impact on the development of any area in the world that is measured by per capita energy consumption on the contribution of the energy sector to the gross domestic product. This study, however, adds to the existing literature on the evaluation of the impact of gas flaring on socio-economic and environment in Nigeria using Niger Delta as case study.

In this paper which is divided into four sections, section one is the introduction, section two is the review of existing literature, section 3 is concerned with the methodology of the study and interpretation of the analysis while section 4 is its concern with the policy implication of findings, recommendations and conclusion.

2.0 A Review of Existing Literatures

The exploration for oil over the years since discovery of first commercial quantity oil in Oloibiri has posed several environmental challenges. Oni and Oyewo (2011) posited that over time, extensive exploration for oil has had a huge negative impact on human health, the local culture and the self aspirations of the people of this region. One of such practices associated with the exploration for crude oil is the flaring of gas into the atmosphere. Wikipedia (2011) describes gas flaring as a process, involves the use of an elevated vertical stack or chimney as a channel through which undesired gas or combustible gas and liquids are burnt as they exit the flare stacks on oil wells or oil rigs.

The origins of gas flaring can be traced to the activities of Shell-BP with the epoch-making discovery of crude oil in commercial quantity at Oloibiri (Bayelsa State) in August 1906 (Jimoh and Aghalino 2000). As Osuoka and Roderick (2005) assert, "the first field was found in 1956 and the first export was made in 1958. Flaring of gas mixed up with crude oil began right at the start, and so did a recognition of its unacceptability" (p.6).

Gas Flaring in Nigeria

At the onset of oil exploration in Nigeria, practice of gas flaring became institutionalized as natural gas was deemed to be a waste product that resulted from the process of exploring crude oil from the ground and the practice became institutionalized throughout the industry (Oni and Oyewo; 2011).

Over the years government after government has formulated policies and measures to try and capture the flared gas

through the setting up of the Bonny LNG project in 1989 together with other gas gathering projects, these measures have been largely adequate to take care of the volume of natural gas been produced.

Apart from operational and bureaucratic and its attendant losses to the economy (Adebayo 2010), gas flaring has constituted environmental nuisance in Niger Delta region of Nigeria.

Nigeria is endowed with a huge gas reserve, in fact, petroleum experts regard Nigeria "as a gas province with little oil in it" (Gaius-Obaseki, 1996). Nigerian gas reserve is estimated to be about 124 trillion cubic feet (TCF) of gas in 2005 which in terms of energy it is said to be twice as much as the nation's crude oil reserves. Natural gas in Nigeria is obtainable in two main forms-Associated natural gas (AG) and as Non-associated natural gas (Non-AG). Approximately 75 percent of the total gas output are flared in 2000. This may be broken down into 8 percent of non-associated gas and 92 percent of the associated gas output (NNPC, 2008). Gas flaring has, thus, become a dominant feature of upstream activity in the petroleum industry of the Nigerian economy (Okoh, 2001).

Gas flaring in Nigeria could be blamed on the unsustainable exploration practices coupled with the lack of gas utilization infrastructure in the country. However, Ojinnaka (2008) believes that energy, such as gas, has a pervasive impact on the economy and environment such that the progress or development of any area in the world is measured by per capita energy consumption or the contribution of the energy sector to the gross domestic product of that area.

As a result of this perception that associated gas was deemed to be a by-product, no facilities were put in place to capture and store the gas.

This was indeed evident at the dawn of exploration in Nigeria when British officials cited economics and a lack of markets for the continued flaring of gas.

Thus, it was cheaper than to flare the gas as against storage. It was felt that the laying of pipelines and the creation of storage tanks was too expensive an undertaking at the time; and coupled with this was the perceived lack of a viable market for the gas. This was, to say the least, double standards on the part of the British in that they took a contrary position in relation to oil exploration and the attendant associated gas flaring in the North Sea, as the practice was totally discouraged from the onset. Thus Nigeria, it seems, was bequeathed with an oil industry that institutionalized the flaring of gas, with the total disregard for the attendant impact on the environment and the huge waste of an energy resource (Osuoka and Roderick 2005). It, thus, becomes inevitable that challenges would exist while trying to re-organize the petroleum sector. These challenges

come in the form of expanding the network of pipelines and increasing the number of gas storage tanks. This coupled with the need to find markets for the gas within and outside the country has resulted in bureaucratic bottlenecks in the form of delays in executing market oriented projects like the West African Gas Pipeline project and the continued somersaults on the part of past and present Nigerian governments as to when to enforce the gas flare-out date on the multinational oil companies. This inaction, on the part of the government and the multinational oil companies can be traced to the fact that the Nigerian oil industry is steeped in corruption; which has been to the benefit of the ruling local elite. As a result, policies of best practice

that should underpin the activities within the industry are non-existent, thus accountability and transparency are words rarely used in the industry and this has had a lasting impact on the continued flaring of gas or indeed any other negative practice within the industry.

Detailed and historical studies have provided empirical validation of the views that the discovery and extraction of natural resources has brought different consequences to countries that are endowed with such resources, while some of these nations have been economically strong, others have been facing serious economic hardships and conflicts.

The studies by Cedigas (2000), Nigeria is by far the number one flarer of natural gas on the planet both absolutely and proportionally - about 46 percent of Africa's total and the most gas flared per tone of oil produced. The data also indicates that Nigeria accounted for 19.79 percent of the global figure.

Orubu (2002b) who undertook a comparison of concentrations of ambient air pollutants in the region and Lagos State concludes that pollutant concentrations are highest in the Niger Delta and argues that some of the green house gases (such as methane and carbon dioxide) emitted at flare sites contributes to global warming.

This suggests therefore that Nigeria oil fields contribute more to global warming through flaring of associated gas than the rest of the world.

The largest proportion of these flare sites are located in the Niger Delta.

An impact assessment of the 1983 Oshika oil spill by Powell and White (1985) confirmed the dearth of floating and submerged aquatic vegetations especially water Lettuce, fish, crabs and birds.

Otukunefor and Biukwu (2005) have all shown the pollution Levels of aquatic ecosystems observed in the region are as a result of unregulated effluent discharges and unsustainable methods of petroleum extraction.

Amakiri (2005) laments the loss of biodiversity, alternation of habitats and deforestation that is associated with petroleum

exploitation related canalization. This canalization which is quite extensive in the region.

The canalization which is quite extensive in the region opens up previously pristine and inaccessible ecosystem to illegal logging activities.

Ndiokwere and Ezele (1990) also report high levels of heavy metals in soils and plants near the Warri refinery.

Enoyan et al (2006a, 2006b) have also confirmed high levels of heavy metal contamination of River Ijana and efficient receiving - steam that flows by the same refinery.

Braide et al, (2004) observed high concentrations of heavy metals in the Niger Delta. Furthermore, Spiff and Horsefall, (2004) reported trace metal contamination of the intertidal flats of the upper New calabar River in the Niger Delta.

Gill et al, (1992) and Agbogidi et al, (2006) conducted an independent studies, that documented adverse effects of crude oil, engine oil and spent lubricating oil on soils and the suppression of germination of seeds as well as stomata abnormalities in diverse food crops.

The major sources of degradation of forest land and water in the region include oil spills, gas leaks, blow outs canalization and discharge of wastes and effluent from oil and gas operations directly into the surface water bodies and the low surface, oil spills in Nigeria was due to a number of causes that include corrosion of pipelines and storage tanks, sabotage and accidents in oil operation.

A world bank survey (Grey, 1995) estimated that about 2.3 million cubic metres of crude oil is split in about 300 separate incidents in the region each year and noted that oil companies normally understate the incidents of spillage and that the total volume of the oil spilt might be as much as ten times the official figure.

The official figure of SPDC (2004) show that between 1976 and 2001, 6,187 incidents in which 3 millions barrels used spilled. Greater than 70 percent of the volume went increased (UNDP, 2006).

Leaks and spills also affect ground water quality. Preliminary results of on going ground water quality evaluation around the CORPC show elevated levels of BTEX in shallow boreholes and dugwell water (Akpoburie et al 2008)

An earlier study by Douglas et al (2005) on the effect of deforestation on tropical basic hydrology has also associated for instance of 1 percent basic yielded for every 3 percent forest loss thereby increasing the probability of flooding.

Efforts at developing the region through various intervention schemes and agencies have been dismal failures and a result of insincerity, dishonesty and lack of commitment on the part of the federal government that initiated the scheme.

Blasing, Hand and Kimberly (2007) worked on monthly carbon emissions from natural gas flaring and cement manufacture in

United State. They discovered that emissions amounted to 1 % of all fossil fuel and carbon emissions had no clear and persistent pattern annually. Akpan (2009) in his work noted that literature shows that vast amount of these gas are being used by few of our chemical industries. He posited further that the only way out for harnessing the nation’s natural gas is to encourage the establishment of gas based petrochemical complexes that can consume large volume needed to eliminate gas flaring. He Abdul Kareem and Odigure (2011) worked on the Economic benefit of Natural Gas utilization in Nigeria based on a case study of the food processing industry their work focused on the measurement of heat radiation from gas flaring as one of the menaces of gas flaring fits of substituting conventional fuel and energy types, such as antomotue gas oil. Results obtained revealed that up to 69% are used, 29.85% on low poor fuel oil and million electricity with natural gas could be saved by the company, translated to millions of dollars in few years if conventional fuel and energy is substituted with natural gas.

4. GAS FLARING ISSUE: AN OVERVIEW

The flaring of gas in Nigeria has resulted to the release of associated gases which have been linked to climate change. The emission of greenhouse gases, in particular, methane and carbon dioxide during flaring continues to induce climate change and its impact on the environment. As Osuoka and Roderick (2005) noted, “The burning of Fossil fuel, mainly coal, oil and greenhouse gases has been inducing global warming and this may get worse during the course of 21st century.

Table 1:
Top 20 flaring countries

S/N	2004 Rank	Reputed..... country
1.	Nigeria	24.1
2.	Russia	14.7
3.	Iran	13.3
4.	Iraq	8.6
5.	Angola	6.8
6.	Qatar	4.5
7.	Algeria	3.7
8.	Venezuela	3.7
9.	Equatorial Guinea	3.6
10.	Indonesia	3.5
11.	USA	2.8
12.	Kazakhstan	2.7
13.	Libya	2.5
14.	Azerbaijan	2.5
15.	Mexico	1.6
16.	UK	1.6
17.	Brazil	1.5
18.	Garbon	1.4
19.	Cameroon	1.1
20.	Canada	1.0
Total Top = 20		107.5
Source: Garet B (2007)		

Table 2: The impact of gas flaring on agriculture output

Distance of farm land	Percentage loss in
From flore site	Yield of crops
200 metres	100 percent
600 metres	45 percent
1 kilometre	10 percent
Source: Salau (1993: 19-23) (Opukri Co. Ibaba IS 2008)	
Adegemo 2000:69	

Figure 1

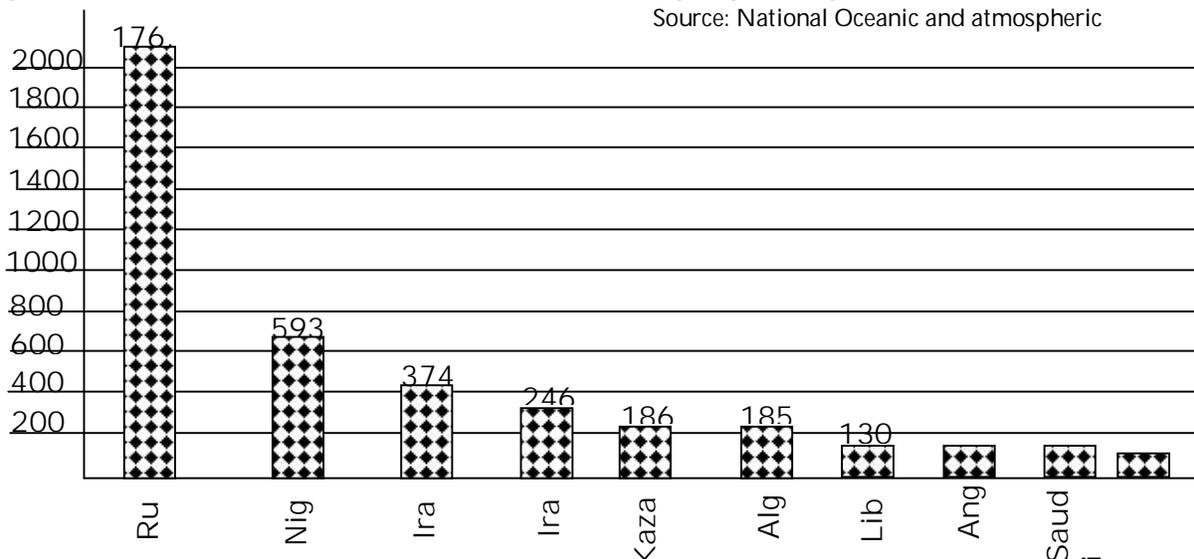


Figure gas flaring 2007
Source: National Oceanic and atmospheric

• Gas Flared			
%gas flared	49.8	41.9	35

Source: Madueme 0 (2%)

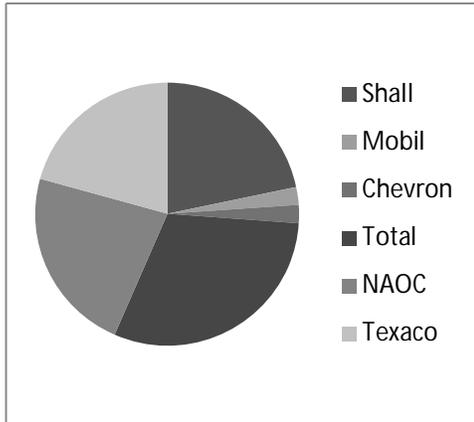
- Billion cubic feet (bef).

Administration in EIA (2009)

Figure:

Total percentage of gas fenced in Nigeria (200-2008)

Source: Madueme 0 (2010)



Source Author: Analysis based on data computing from table 4

Table 3: Major oil companies with data on natural gas produced and flared for the period (2000-2008)

Name of company	2000-2002	2003-2005	2006	2007	2008
Shell					
• Gas produced	1632857255	2114726414	735315496	783905871	800689882
• Gas flared	835756839	784786431	16340866	90612316	10715991
• Mobil					
• Gas produced	12323233154	1159585960	49110702	464537132	130586764
• Gas flared	383175395	535622854	210026922	183528046	
Chevron					
• Gas Produced			235249063	655500024	243040550
• Gas Flared	627587155	389895189	192602299	101186784	156952278
Total exp	455440050		(1)	162780356	
• Gas Produced			2180439	802855214	202360
• Gas Flared	332525971	802855214	2180439	802855214	202360
• NAOC					
• Gas Produced			6422492	265047485	202360
• Gas Flared	123648616	265047485	6422492	265047485	202360
• TEXACO					
• Gas Produced	1109643063	1244207143	423716209	33849981	293668636
• Gas Flared	587136497	496718413	109926431	320927714	96353534
• TEXACO					
• Gas Produced	87486761	3702939	5941278	2479303	4803727
• Gas Flared	87031562	36653106	5828277	2479303	4746874
Pan Ocean					
• Gas Produced	62940758	74517198	3941639	21752432	21752432
• Gas Flared	59828073	70969973	3756324	0.00	21211546
Ground Totals	50855364117	6088397892	2114245718	0.00	2111442905
• Gas Produced	2532017032	2549693811	740770521	2032853975	542732471

5.0 METHODOLOGY AND EMPIRICAL ANALYSIS

Data were obtained from a community survey (2011) of information on gas flaring issues and its environmental impact on Nigeria Delta Community in Nigeria. The survey which include five communities in three local Governments of both Bayelsa and Delta State. The five communities are Oghulaha, Odimodi, Ogbotubo, Arotan and Sagara are used as case study.

Two hundred and fifty questionnaires were distributed randomly in the communities and collected back within (3) days out of the returned questionnaires two hundred questionnaires were verified fit for the analyses. The study was able to collect information on Air flaring, drinking water, health status, source of income culture, religion and Aesthetic from three local Governments of five communities in two states. Hundred percent of the final data used was obtained from the communities survey in both Bayelsa and Delta State.

The study adopted the model of Kareem and Odigare (2011). The data sourced from the Central Bank of Nigeria statistical bulletin were using ordinary least square technique while the information obtained from the community surveyed were analyzed with the use of chi-square technique. The dependent variable is economic growth measured as Gross Domestic Product and others are explanatory variables are volume of gas flared (VGF), crude oil production, gas flaring rate etc.

The model can be written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \mu$$

Where X_1 is the volume of gas flared, X_2 is the amount crude oil production, X_3 is the gas, flaring rate and X_4 is the volume of gas produced and μ is the stochastic error term.

Equation 1 was estimated using an ordinary least square (OLS) estimator. The OLS estimator has a large number of desirable properties. This actually makes it the best linear unbiased estimator.

Table 4 RESULT AND ANALYSIS

It was discovered that there is a relationship between economic growth (GDP) crude oil production (COP) and gas flaring (GF) hence it is concluded that natural gas flaring is positively related to the level of demand for petroleum product. It was also learnt that demand for crude oil international triggered production of natural gas as it is established that associated

Gas formed the bulk of Natural gas production in Nigeria. It was also discovered that there exist a relationship between volume of Gas flaring fine (GF), gas flaring fine (GFF) and volume of Gas produce it is established that fines impose on gas – flaring does not influence the level of gas flared in Niger Delta.

The hypothesis highlighted from the information obtained from the community survey is stated as :Ho: That the activities of the oil companies in Niger Delta region does not have any adverse effect on people of Niger Delta communities.

The chi-square formula is stated as $X^2 = \frac{\sum(O-E)^2}{E}$.

Where $\sum O$ is the summation, O is the observed frequency and E is the expected frequency. X^2 cal-refers to computation for the test statistics and X^2 table value .The degree of freedom (df)=(R-1)(C-1) where n=5 is 4.

At 5% level of significance ,the degree of freedom is 4 and X^2 cal=617.43 and X^2 tab=9.49

The decision rule is that if X^2 cal is greater than X^2 tab at 0.05 level of significance, the null hypothesis is hereby rejected and concluded that the activities of the oil companies in the Niger Delta region have an adverse effect on people of Niger Delta.

6.0 CONCLUSION AND RECOMMENDATION

This study has examined the impact of gas flaring and power plant emission on the socio economic environment of the Niger Delta people in Nigeria. The results reveal that gas utilization has significant impact on the economy and it is also sustainable. It reveals further that since the imposition of fine on flared gas in 1984, no structural change has been observed.

Therefore, there is an urgent need for the government to provide environment that is conducive for investment in the gas industry as this will lead to additional income to both the people and the government of Nigeria.

These results also show that the imposition of fine on flared gas may not be a better policy option that the need to provide facilities that will enhance further utilization of Nigerian Natural gas.

Polices

Based on our results and their policy implications.

REFERENCES

- Adeyemi AS 2000. Climate Change and Environmental Threat. In: HI Jimoh, IP Ifabiyi (Eds.): *Contemporary Issues in Environmental Studies*. Ilorin: Haytee Press, pp. 158-164.
- Aghalino SO 2000. Petroleum Exploitation and Environmental Degradation in Nigeria. In: HI Jimoh, IP Ifabiyi (Eds.): *Contemporary Issues in Environmental Studies*. Ilorin: Haytee Press, pp. 141-147.
- Alakpoda JJ 2000. Soil Characteristics Under Gas Flares in the Niger-Delta, Southern Nigeria. *Geo-Studies Forum*, 1(1 and 2): 1-9.
- Aghalino SO 2000. Petroleum Exploitation and the Agitation for Compensation by Oil Mineral Producing Communities in Nigeria. *Geo-Studies Forum*, 1(1, 2): 11-20.
- Akasike C, Adelakun A 2010. Back from the kidnappers' den. *The Punch*, July 19, 2010, P.3.
- Ebiri K 2010. How govt can gain N'Delta people's trust, by MOSOP. *The Guardian*, June 24, 2010, P.80.
- Energy Information Administration. 2009 Country Analysis Briefs: Nigeria. <http://www.eia.doe.gov> (Retrieved January 28, 2010)
- Garvet B 2007. 'Gas Flaring Emissions Contributes to Global Warming.' http://www.itu.se/polopoly_fs (Retrieved July 12, 2010)
- Ikporkpo CO 1983. Environmental Deterioration and Public Policy in Nigeria. *Applied Geography*, 3: 303-316.
- Ishisone M 2010. Gas Flaring in the Niger Delta: The Potential Benefits of its Reduction on the Local Economy and Environment. <http://nature.berkley.edu/classes/es196/project>. (Retrieved June 28, 2010)
- Jimoh HI, Aghalino SO 2000. Petroleum and Environmental Degradation: A Perspective on Government Policies in Nigeria. In: HI Jimoh, IP Ifabiyi (Eds.): *Contemporary Issues in Environmental Studies*. Ilorin: Haytee Press, pp. 238-244.
- Madueme S 2010. Gas flaring activities of major oil companies in Nigeria: An economic investigation. *International Journal of Engineering and Technology*, 2(4): 610-617.
- Madueme S 2010. Economic Analysis of Wastages in the Nigerian Gas Industry. *International Journal of Engineering and Technology*, 2 (4): 618-624.
- Malumfashi GI 2008. Phase-Out of Gas Flaring in Nigeria by 2008: *The Review of the Regulatory, Environmental and Socio-Economic Issues*. University of Dundee, Scotland. Prospect of a Multi-Win Project.
- Osunde A 2009. Gas Flaring Cost Nigeria \$1.46bn Annually. *Reuters*, March 11, 2009, P.1.
- Opukri CO, Ibaba IS 2008. Oil Induced Environmental Degradation and Internal Population Displacement in the Nigeria's Niger Delta. *Journal of Sustainable Development in Africa*, 10(1): 173-193.
- Osuoka A, Roderick P 2005. Gas Flaring in Nigeria: A Human Rights, Environmental and Economic Monstrosity. *A joint report written by the Environmental Rights Action/ Friends of the Earth, Nigeria and The Climate Justice Programme*, Amsterdam: Press.
- The Economist 2009: http://www.economist.com/displayStory.cfm?story_id=1097980 (Retrieved May 3, 2010)
- Wikipedia 2009 Gas Flaring. http://en.wikipedia.org/wiki/Gas_flare (Retrieved May 3, 2010).

Appendix

Log likelihood - Durbin-Watson stat 1.365962 256.85823
77258 8413 9127

E. View Results

Model I

Dependent Variable; GDP

Method; Least Square

Date: 09/02/22 Time: 17:43

Sample: 1982 2006

Included Observations: 25

GDP=C(1)+C(2)*VGF+C(3)*COP

C(1)	Coefficient	Std. Error	t-Statistic	Prob.
		-4000332.42		-0.001703
	14288505	137	3.57182961	20425169
		7939	036	
C(2)	-66.3810550	-0.898827		
	8.5378902	148	0.12861938	405328
		6936	1953	
C(3)	26.411730	6.16259973	4.28580980	0.000300
	346	342	244	18007021

7

R-squared	0.4803889	Mean dependent var	3398346.
	21425		252
Adjusted R-square	0.4331515	S.D depended var	4982461
	50645		16674
S.E of regression	3751262.0	Akaike info criterion	33.22524
	5404		9473
Sum square resid	3.0958327	Schwarz criterion	33.37151
	3958e+14		4572
Log likelihood	- Durbin-	Watson stat	0.306951
	412.31561		8413

Model II

Dependent Variable; VGF

Method; Least Square

Date: 09/02/22 Time: 17:43

Sample: 1982 2006

Included Observations: 25

GDP=C(1)+C(2)*VGF+C(3)*COP

C(1)	Coefficient	Std. Error	t-Statistic	Prob.
	-4552.1360	5421.41968	0.83965756	0.410132
	6359	438	7317	568023
C(2)	783.54106	5821.90298	0.13458504	0.894163
	5136	171	3344	665345
C(3)	0.4835392	0.07889586	6.12882841	3.612307
	12168	38469	496	95893e-

R-squared	0.6591312	Mean dependent var	23255.79
	54836		2
Adjusted R-square	0.6281431	S.D dependent var	12257.90
	87094		14947
S.E of regression	7474.9126	Akaike info criterion	20.78865
	8782		91302
Sum square resid.	122923503	Schwarz criterion	20.93492
	3.19		42291

Juridical Mass for Protection of Waters from Pollutants – Heavy Metals in Water of Lake Ohrid and in Fish Species – *leuciscus cephallus albus*

Agni Aliu, Suzana Aliu

Abstract— Metals occur in different forms: as ions dissolved in water, as vapours, or as salts or minerals in rock, sand and soil. They can also be bound in organic or inorganic molecules, or attached to particles in the air. Both natural anthropogenic processes and sources emit metals into air and water. The effects of metals in the environment depend to a large extent on whether they occur in forms that can be taken up by plants or animals. Uptake of metals in an animal involves metal ions crossing a cell membrane.

The aim of this paper was analysis and determination of heavy metals like Cu, Zn, Cd and Pb in water samples of different profiles from Ohrid Lake. Studied water samples were taken in different locations and different depths (0m, 20m, 40m, 75m, 100m and 150m). The presence of toxic heavy metals in food chain was studied in fish species – Klein (*Leuciscus Cephallus albus* Bp Bonaparte). Determination of analysed heavy metals was done using Atomic Absorption spectroscopy using a Perkin Elmer 370 A and 370 flames – aer acetylene and AAS Peyunicam 926 model.

Index Terms— Bones, heavy metals, gills, liver, muscular tissues.

1 INTRODUCTION

Metals occur naturally in the environment and are present in rocks, soil, plants, and animals. Often a ligand, or a carriers, executes this transport. Sometimes there are additional specific carriers within the cell. If an organism's uptake of metal is greater than its ability to get rid of it, the metal will accumulate. [2.4. 13]

Heavy metals tend to accumulate in storage compartments. For example, cadmium accumulates preferentially in the kidneys, and lead in the skeleton. The accumulation can continue throughout the organism's life and is the major cause of chronic toxicity. In contrast to organic pollutants, metals accumulate in protein tissues and bone rather than fat. [3,5]

2 OHRID LAKE

Lake Ohrid (Macedonian: transliterated: *Ohridsko Ezero*; Albanian: *Liqeni i Ohrit*). It is one of Europe's deepest and according to most experts the oldest lake in Europe, preserving a unique aquatic ecosystem with more than 200 endemic species that is of worldwide importance. The importance of the lake was further emphasized when it was declared a World Heritage site by UNESCO in 1979. [9.12]

However, human activity on the lake shores and in its catchment area is resulting in the ecosystem coming under stress. Lake Ohrid is situated in the Southeastern part of Europe, with Albania and FYR of Macedonia as the only two lakeshore states. It is peculiar for its rich biodiversity, that derives of its very old age. Ohrid Lake is the biggest and the deepest from all the lakes of the group known as "Desaret". It is called a Museum of living fossils and it offers a living environment for too many organisms typical for mild waters, which, in the Balkans Peninsula and Central Europe are found only in the fossil form. [15].

Another typical peculiarity of this Lake is its the only flow - the river Drini i Zi, with an approximate flow of 22.4 m³ s, for most of the time.

3 LIVING SPECIES

Ohrid Lake is well known for its 17 kinds of fish. They are all autochthonous species. Ten of them are commercially important. [7.8.11]

Another peculiar feature of this lake is its ability to self regeneration. Its water comes from different sources as well as from its water accumulating basin.

4 SOCIO-JURIDICAL ASPECTS FOR PROTECTION OF LIVING ENVIRONMENT

Protection of living environment is one of the most important fields of our time and, it is more than a logical thing that, the society should show special attention to this field in R. of Macedonia, as well. [1]

This is stated in act 8 of Constitution of R. of Macedonia where arrangement and space humanity and protection and

-
- Doc. Dr. Angi Aliu, Faculty of Public Administration and Political Science - South East European University, Tetovo- Macedonia
 - PhD. Dr. Suzana Aliu, Faculty of Food Technology and Health - State University of Tetovo - Macedonia

advancement of living environmental and nature and respect of the norms of international Law, in general is accepted and is stated as one of the fundamental amount of Constitution of R. of Macedonia [6,10,14].

Clean Living Environment should be guaranteed to every citizen. Everyone is obliged to protect living environment and nature that means to protect the soil, the air and water.

Of course, regulation of living environment and nature protection, and it need to be arranged with other special laws that must have accordance with international norms and Constitution of R. of Macedonia. But in a state not always inside law's and International Convents are observed, like changing riverbed of river Sateska in 1962. Before this year this river was pour in River Drini i Zi and without a reason its riverbed diverted and now it pours in Lake Ohrid, even if a thing like this is disallowed with International Konvent and with protection law of Lake Ohrid, Prespa and Dojran (act 7) [1,6].

Along pouring this river brings with itself dangerous materials with which humans and other organisms are in dangerous. This materials can bring change's on water quality and world life, in physical-chemical, biological, radiological, microbiological aspects, etc.

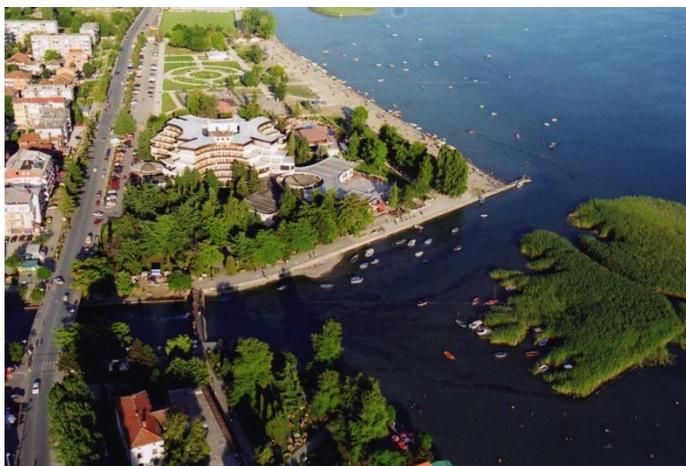


Fig. 1. Images of a beautifull part of Lake Ohrid



Fig. 2. Images of dirty parts of Lake Ohrid

5 EXPERIMENTAL ENGAGEMENT MATERIAL GATHERING AND WORKING METHODS

In this output, the presence of heavy metals (Fe, Mn, Cu, Cd, Pb, Ni, Zn) in the water on its vertical profile was analyzed, beginning from levels (0m, 20m, 40m, 75m, 100m and 150m depth), as well as their impact in the nutrition chain. Three exemplars of fish known as *Leuciscus Cephallus Albus*, L1, L2 and L3 were examined. From each of them organs like gills, liver, muscular tissues and bones were dissected.

After the dissection, the determination of the concentration of the heavy metals, such as: Zn, Cu, Cd and Pb in each of the organs was done.

To achieve this, SAA - Spectroscopy of Atomic Absorption - the flame technique was used. The measurements were done by Spectrophotometer of the atomic absorption of the type PERKIN - ELMER 370 A and 370. Flame - Air - Acetylene and SAA PEYUNICAM 926. (4)



Fig. 3. *Leuciscus cephallus albus*

6 EXPERIMENTAL RESULTS

Achieved results are presented in charts / graphics as follows: Charts 1-5 to represent the achieved results of the concentration of the heavy metals on the studied samples through experimental work.

TABLE 1
DETERMINATION OF HEAVY METALS IN THE WATER ON THE VERTICAL PROFILE OF THE LAKE BY SAA- MG/L

Sample	L1	L2	L3
Zn	25.1	10.3	19.9
Cu	14.8	214.6	27.3
Cd	0.388	2.691	0.691
Pb	-	-	-
gills	2.956	0.135	0.353
Liver	1.745	0.125	0.145
Zn	0.91	2.64	2.69
Muscular tissues	3411	5312	2217
Bones	0.14	0.49	1.1
Pb	0.415	0.072	0.075
Pb	0.09	-	-

TABLE 2

PARAMETERS OF ANALYSED FISH

Sample	L1	L2	L3
Zn	2.71	186.8	3.33
Cu	10.2	227.8	77.4
Cd	0.14	1.18	0.46
Pb	3.29	-	4.01

TABLE 3
 DETERMINATION OF HEAVY METALS ON GILLS OF FISH WITH SAA – µG/L

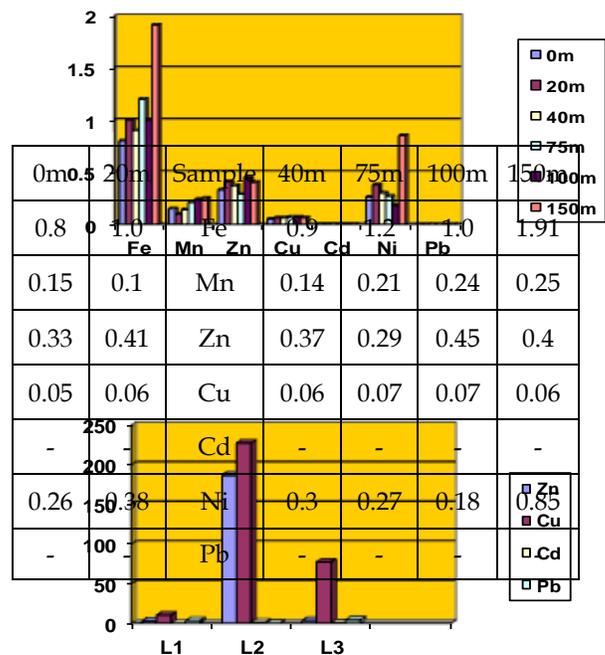
TABLE 4
 DETERMINATION OF HEAVY METALS ON LIVER OF FISH WITH SAA – µ µg/l

TABLE 5
 DETERMINATION OF HEAVY METALS ON THE FISH MUSCULAR TISSUE WITH SAA – µ µg/l

TABLE 6
 DETERMINATION OF HEAVY METALS ON FISH BONES WITH SAA – µ µg/l

Sample	L1	L2	L3
Zn	7.85	11.5	7.73
Cu	46.4	266.9	412.8
Cd	0.34	3.52	2.4
Pb	0.91	2.64	2.69

Graphic charts from A to E represent the same results – the results from the previous charts 1-5, added is only a better visual representation of them.



7 DISSCUSION OF THE RESULTLS

Fig. 4. Chart A - Concentration of Fe, Mn, Zn, Cu, Cd, Ni and Pb in water on the vertical profile of the lake, determined with SAA (mg/l).

Fig. 5. Chart.B - Concentrtion of Zn in the organs of fish, determined with SAA ($\mu\text{g/g}$)

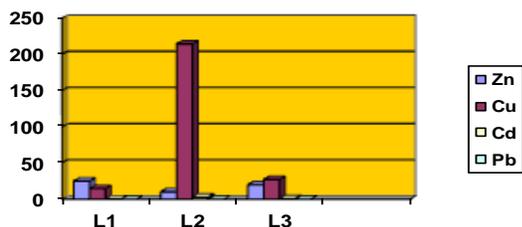


Fig. 6. Chart C - Concentration of Cu in the organs of fish, determined SAA ($\mu\text{g/g}$)

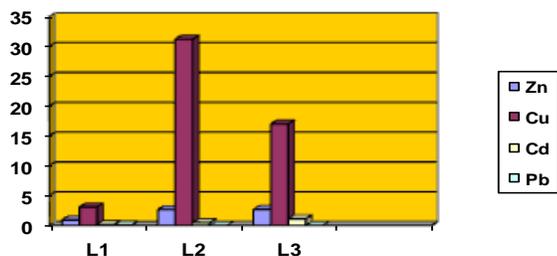


Fig. 7. Chart D - Concentration of Pb in the organs of fish, determined SAA ($\mu\text{g/g}$)

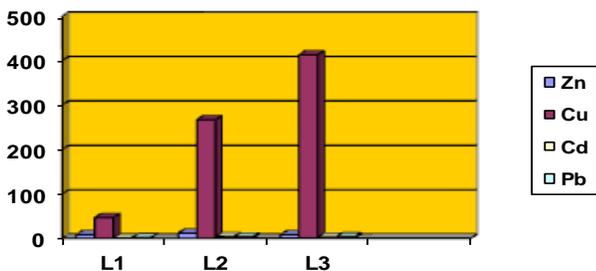


Fig. 8. Chart E - Concentration of cadmium in the organs of fish, determined SAA ($\mu\text{g/g}$)

After having thrown an analytic view on the results given above through the charts, both numeric and graphic ones, one can notice very clearly that on the vertical profile of the lake waters, the concentration of the metals were traced, as follows:

Concentration of Mn goes from 0.10 mg/l in 20 m to 1.91 mg/l in 150 m deepness (Tab.1,Chart A).

Iron - Fe was represented in concentrations 1.0 mg/l in 20 m up to 1.91 mg/l in 150 m deepness. (Tab.1,Chart A)

Zinc was registered on minimal values from 0.41 mg/l in 20 m, up to 0.45 mg/l in 100 m deepness. (Tab.1,Chart A) Copper was registered from 0.07 mg/l in 20 m to 0.07 mg/l in 100 m deepness. (Tab.1,Chart A)

Nickel gave these results: 0.38 mg/l in 20 m up to 0.85 mg/l in 150 m deepness. (Tab.1,Chart A). From the analyses done on the studied water samples, about the concentration of the above mentioned metals, one can easily understand that, with some exceptions, their concentration on the lake waters is not of a high level. That's why, we can say that the waters on the vertical profile of the Ohrid Lake belong to the waters of the first or second class.

In order to have a realistic view about what was stated above, during the determination of the presence of the heavy metals in the organs of fish, our results have shown the values given below:

In order to have a realistic view about what was stated above, during the determination of the presence of the heavy metals in the organs of fish, our results have shown the values given below:

Minimal amount of Zinc was accumulated in the muscles 0.91 $\mu\text{g/g}$ L1 (Tab 4, Chart B), up to 186.8 $\mu\text{g/g}$ in the gills L2 (Tab 2, Chart B).

Copper did give these results: 3.1 $\mu\text{g/g}$ - L1 in muscles (Tab 4, Chart C) up to 412 $\mu\text{g/g}$ L3 in the bones (Tab 5, Chart C).

The amount of the concentrated Lead goes from 0.09 $\mu\text{g/g}$ in muscles - L1 (Tab 4, Chart D), up to 4.01 $\mu\text{g/g}$ in gills - L1 (Tab 2, Chart D).

Results of the Cadmium have shown that the gill of L3 possesses a concentration of 0.14 $\mu\text{g/g}$ in muscles - L1 (Tab 4, Chart E), up to 3.52 $\mu\text{g/g}$ in the bones of L2 (Tab 5, Chart E).

4 CONCLUSION

Presence of the heavy metals in the leaving creatures is done by their penetration through the cell membrane. This process is frequently done through some conductors specific for the cells.

Metals get accumulated in the organism in cases when the conductivity of a given metal is bigger than the capacity of the organism to get rid of it.

It is typical for the heavy metals to get themselves accumulated in the accumulative organs. This way, cadmium initially is accumulated in the kidney, copper in liver and Lead in skeleton.

This process goes on during the whole lifetime of the organism and occasionally it brings it to the level of chronic poisoning. From the above results, one can notice that the fish belbica does accumulate a higher amount of heavy metals and this amount get bigger together with the weight of the fish.

From what was stated above, one can conclude that the accumulation of heavy metals by the leaving creatures that do live within water ecosystems, is an ongoing process that is not preferred because it represents a sort of threat for their mere existence as well as for the nutrition chain to come.

Using of chemicals (pesticides) must be controlled in agriculture especially along watershed of the river wich are pure in the Lake Ohrid.

It is necessary bringing and application of rigorous juridical dispositives with international and national character. Prevention measure's to delay river's and lakes pollution, which pour in Lake Ohrid.

International co-operation and in local degree between Macedonia and Albania.

Growth of the degree of cooperation between governmental organizations and nongovernmental organizates.

Cooperation of organs of central government (ministry) with organs of local government (sector for living environment protection) and co-operation of units of local self-government (Commune) between them.

REFERENCES

- [1] A.Aliu; The decentralisation Process and the Local Selfgovernment in R.Macedonia after the Ohrid Framework Agreement; Tetovo; 2008;pp.
- [2] Balvay; What is the best time reference to express biological results.The calendar year or the ecological year.Verh.Internat Verein. Limnol. (1991)24:928-930.
- [3] C.R.Goldman; Home,Limnology.Mc.Graw-Hill Book Company, New York St. Luis-San Francisco-Auckland. Bogota- Hamburg-Toronto.1983.pp.464.
- [4] CM .Arsoski; Nekoi pogledi vrz genezata na Ohridskoto Ezero. MANU. Simpozium za problematikata na regulacijata na Ohridskoto Ezero. 1974. Pp.76-85.
- [5] D. Twenddle; Conservation and threats to the resources of Lakes Malawi. Mitt. Internat. Verein. Limnol. 1992. Vol. 23 pp. 17 – 24
- [6] Constitution of Republic of Macedonia; Tetovo 2007; pp.11,20
- [7] Ernst Basler and Partner 1995. The World Bank; Feasibility Study on the Lake Ohrid Conservation Project. Zollikon.
- [8] J.Hadište; Dosegašnite sovetovanja za problemite na zaštitatata na Ohridskoto Ezero so negovata okolina i novi predlozi za nivno rešavanje. DEM. Zbornik na trudovi "Ohridskoto Ezero i negovata zaštita. 1992. pp. 19 – 26.
- [9] K. Risteski; B.Davitkovski; Ekološko pravo, Skopje, 1997, pp18-22.
- [10] Low for protection of Lakes Ohrid, Prespa and Dojran. Skopje, sluzben vesnik br. 62/93.pp.3.
- [11] Mitrevski,K.,D,Davvcev,M,Jordanoski,andT,Naumoski.1997.Pollution monitoring system for the Lake Ohrid- a planning scheme , Water pollution IV : Modeling, Measuring and Prediction, R. Rajar and C.A.Brebbia,eds.,pp.341-347
- [12] R. Hecky; The eutrofication of Lake Victoria. Verh. Internat. Verein. Limnol. 1993. Vol. 25 pp. 39 – 48.
- [13] R. Beckett; Surface and Colloid Chemistry in Natural Waters and Water treatment, Plenum Press, Melbourne. 1990.
- [14] S.Aliu; Environmental pollutions Lake Ohrid; Lambert Academic Publishing Book -Seebrucken -Germany, 2010 pp. 15 - 16.
- [15] T. Naumovski; Fosforno optovaruvanje na Ohridskoto Ezero. Dokt. Diser. Univ. "Sv. Kiril i Metodij" PMF – Institut za biologija ; Skopje. 2000. pp. 1 – 37.

A New Approach to Mechanochemically Synthesizing Al₂O₃/Cu-Cr Nanocomposites

Elahe Hosseini Nezhad, Niloofar Heidari, Mohammadali Ghorbani, and Alireza Heidari

Abstract—In this study, through the mechanochemical synthesis, the nanocomposite Al₂O₃/Cu-Cr was fabricated from a primary mixture of CuO, Al, Cu, and Cr₂O₃ with different percentages by means of the energetic milling. The fabrication of this nanocomposite was based on the regenerative reactions. Such that, the chromium oxide and the copper oxide are regenerated by the aluminum oxide during the process and the final product is a metal matrix composite including the supersaturated chromium solution in copper as the base and the aluminum oxide as the booster phase. Using the extra metallic copper in the primary mixture as the diluent provides the possibility of controlling the adiabatic temperature and the reaction rate. The results of the computation of the lattice parameter based on the XRD data indicate that, for the copper, this parameter changes during the milling process. These changes in the lattice constant are related to the dissolution phenomenon of the chromium and the formation of the supersaturated chromium solution in the copper. Also, the results of the XRD show that, as the primary composition of the powdery mixture and consequently, the adiabatic temperature of the tested samples are changed, the progress of the reactions rate changes. The SEM results demonstrate that, over a 30-to-48-hour period of milling, the fragmentation phenomenon of particles (the shrinking of particles) predominates in comparison to the surface welding of particles (the agglomeration phenomenon of particles). After 48 hours of milling, a nanocomposite containing particles with the approximate size of 30 nm is achieved.

Index Terms—Al₂O₃/Cu-Cr, energetic milling, mechanochemical synthesis, nanocomposite, regenerative reactions

1 INTRODUCTION

Thanks to the abundant applications in the engineering materials, the metal-ceramic composites have been widely investigated by researchers in recent years. The metal-metal oxide composites also account for a subcategory of the metal-ceramic composites. Al₂O₃ is widely used in the fabrication of the metal-metal oxide composites owing to its low price, high chemical stability, high resistance against oxidation, and high temperature resistance that Al₂O₃/Cu-Cr is one of those composites. This composite with the metallic base of copper-chromium, in which the hard particles of Al₂O₃ are dispersed, possesses the excellent thermal and electrical conductivity of Cu-Cr and the strength, the chemical and the thermal stability of Al₂O₃ together. The Al₂O₃/Cu-Cr composite has been widely applied on account of its very high strength in high temperatures, abrasion resistance, and desirable electrical properties [1], [2], [3].

One of the essential and main aspects of the fabrication process of these composites is that, the oxide particles should be miniaturized as far as possible and dispersed uniformly in the metallic base. This necessity cannot be fulfilled completely through the conventional melting and casting methods. Hence, the powder metallurgy is utilized to fabricate these composites. The first step in fabricating this category of composites through the powder metallurgy is the preparation of composite powder. The composite powders containing oxide particles can be fabricated through the internal

oxidation method, but the composite powder Al₂O₃/Cu-Cr fabricated by this method is not uniform and the volume percent of Al₂O₃ is limited. Therefore, the flash freezing and the mechanochemical synthesis are proposed to fabricate this kind of composite powders [3], [4]. However, the classical application of the mechanical alloying is to fabricate the alloys from the basic elements, the several reports show that, by virtue of the mechanical alloying, a wide range of chemical reactions can be performed. Therefore, many researchers have successfully employed the mechanical alloying to activate the oxidation-regeneration reactions including a proper oxidizer and also a proper regenerator [4]. The fabrication of the composite Al₂O₃/Cu-Cr through the mechanochemical synthesis is based on the below reaction:



The product of the above reaction will be a composite with the copper-chromium base and the scattered booster particles of Al₂O₃. With regards to the huge difference in the heat of formation of Al₂O₃, CuO, and Cr₂O₃ oxides, the reaction among the active metals of Al, CuO, and Cr₂O₃ is highly pyrogenic and categorized as the self-propagating high-temperature synthesis (SHS). The various reports indicate that, in some cases, this regenerative reaction in the mechanical alloying system has been performed spontaneously and combustively or mechanically induced self-propagating reaction (MSR). Rising the temperature during the milling process near to the adiabatic temperature of the reaction causes the Al₂O₃ particles to melt and; consequently, to become coarse which affects the mechanical and physical properties of the composite [3], [5]. Hence, many studies have been conducted in order to perceive the governing mechanisms of this reaction [6], [7] and to achieve a

- Elahe Hosseini Nezhad, Institute for Advanced Studies, Tehran 14456-63543, Iran.
- Niloofar Heidari, Institute for Advanced Studies, Tehran 14456-63543, Iran.
- Mohammadali Ghorbani, Institute for Advanced Studies, Tehran 14456-63543, Iran. E-mail: mohammadalighorbani1983@yahoo.com
- Alireza Heidari, Institute for Advanced Studies, Tehran 14456-63543, Iran.

method for controlling the chemical reaction among Al, CuO, and Cr₂O₃ [3], [8].

The purpose of this study is to investigate the possibility of fabricating these composites through the energetic milling, and the Merzhanov criterion is utilized to predict how the reactions perform. In this study, the mixture of Al-Cr₂O₃-CuO-Cu is used and the composites Al₂O₃-Cr-Cu containing different volume percentages of the Al₂O₃ particles (16-23%) are fabricated through the mechanochemical synthesis. Using different values of metallic copper in the above mixture provides the possibility for controlling the adiabatic temperature, and consequently, for controlling the size of the oxide particles. Moreover, by means of this method, the proportion of the oxide phase dispersed in the base will be controllable. The properties of the fabricated composite powders are investigated by the scanning electron microscope (SEM) and the X-ray diffraction analysis (XRD).

2 THERMODYNAMIC INVESTIGATION OF THE SYSTEM

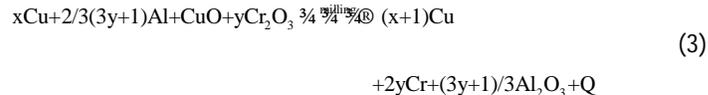
The regenerative substitution reactions as the general form of:



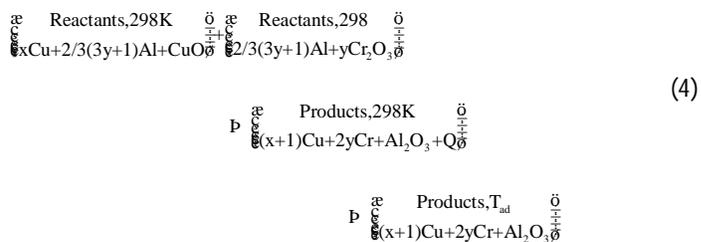
are abundantly applied to prepare the engineering materials and especially the metal-ceramic composites. In this category, one of the important reactions is the thermite reaction during which the oxide of a less active metal is regenerated by Al and forms Al₂O₃. If there is an above average difference between the enthalpies of formation of the metal oxide and Al₂O₃, the regenerative substitution reaction will be pyrogenic and categorized as the SHS reactions. If the required activation energy to reach the ignition temperature (T_{ig}), which is necessary to initiate the reaction, can be provided through the milling process, the reaction is classified as the MSR. In order to categorize easier this type of reactions, the Merzhanov criterion (T_{ad}>1800 K) is used, which is mainly used for the SHS reactions. With regards to the Merzhanov criterion, the reactions whose adiabatic temperature is more than 1800 K can be performed through the MSR method. It is worth mentioning that, there is also a same criterion which introduces a simpler prerequisite ((ΔH₂₉₈/C₂₉₈)>2000 K) to perform the reactions according to the MSR. In the second criterion, the high heat capacity of the reaction's components is not considered. This criterion gives a general overview of the initial moment of the reaction [9].

The results of the studies on the Al-Cr₂O₃-CuO system with the energetic milling SPEX demonstrate that, one to two minutes later the beginning of the milling process, the combustion is performed and the heat released is such that, leads to the evaporation of an amount of copper and the melt of Al₂O₃, and makes the reaction difficult to be controlled during the milling process [5]. To solve this problem, in the system considered in this study, different values of the metallic copper as the diluent and the reaction controller are added to the primary composition and the composite Al₂O₃-

Cr-Cu is fabricated during the milling process and according to the below reaction:



The metallic copper added to the above system increases the heat capacity of the powdery composition; consequently, decreases the temperature rise caused by the combustion. Also, with reducing the direct contact between the particles of CuO, Cr₂O₃, and Al, it controls the reaction rate. Certainly, should be noted that, adding too much copper decreases the system's tendency towards the MSR. In order to investigate the impact of adding copper, the system's adiabatic temperature for different values of the extra copper is calculated according to the below equation, and based on the Merzhanov criterion, the borderline of the MSR reaction is determined. The computation of the system's adiabatic temperature is performed in accordance with the below illustrated path:



Since the reaction is assumed adiabatic, it can be written:

$$\Delta Q = \Delta H_{298}^{\circ}(\text{reaction}) + \int_{298}^{T_{ad}} C_p(\text{products}) \cdot dT = 0 \quad (5)$$

Table 1 indicates the computed values of the adiabatic temperature according to the extra mole of copper. With the addition of the copper to the primary mixture, the adiabatic temperature can be increased through preheating the primary mixture. Therefore, if a reaction is not within the range of the MSR T_{ad}>1800 K, the possibility of performing the process as the MSR can be provided by preheating the primary composition.

3 MATERIALS AND RESEARCH METHOD

The materials used in this study are produced by the powder metallurgy company Institute for Advanced Studies (AIS), and include the copper powder with a purity of 99.78%, the aluminum powder with a purity of 99.15%, and the copper and chromium oxide powders with a purity of more than 99%. The total weight of the powdery mixture containing Cu-Al-Cr₂O₃-CuO is 20 g in each test. The samples S₁ to S₄ are milled for the time periods of 5, 10, 20, and 48 hours until the regenerative reactions are completed (certainly, providing that

the regenerative reaction is completed, the milling process is stopped). The milling process is performed by the 24 stainless steel balls with the diameters of 8 to 16 mm in a closed steel milling container. The ball-to-powder weight ratio is 20:1 and the milling process is performed in a continuous energetic mill. After milling, the samples are investigated by the SEM microscope equipped with the EDX and the XRD phase analysis. The composition of the primary mixture and the composition of the final product after completing the reaction in the tested samples are indicated in Table 1.

TABLE 1										
THE PRIMARY AND FABRICATED COMPOSITIONS (WEIGHT PERCENT) AND THE CALCULATED ADIABATIC TEMPERATURE										
$X\text{Cu} + \text{CuO} + y\text{Cr}_2\text{O}_3 + 2/3(3y+1)\text{Al} \rightarrow (3y+1)/3 \text{Al}_2\text{O}_3 + (X+1)\text{Cu} + 2y\text{Cr}$										
Sample	X	Y	Primary composition (%Wt)			Synthesized composition (%Wt)				T_{ad} (K)
			%Cu	%CuO	%Cr ₂ O ₃	%Al	%Al ₂ O ₃	%Cu	%Cr	
S ₁	1.5	1	24	20	38	18	34	40	26	2034
S ₂	2.18	0.283	47	27	15	11	21	69	10	2480
S ₃	0	1	0	26	50	24	45	21	34	2877
S ₄	6	0.2	73	15	6	6	10	86	4	1604

4 RESULTS AND DISCUSSION

The X-ray diffraction pattern relating to the samples with the primary compositions of S₁, S₂, S₃, and S₄ with different time periods of milling are illustrated in Fig. 1. In this study, considering that the investigation of the reactions' progress during the milling process is based on the XRD's results, the XRD result of the S₁ to S₄ samples is indicated in Fig. 1. By means of investigating these results, it can be observed that, in the samples of S₁ and S₄ after 10 hours of milling and in the sample S₂ after 48 hours of milling, the regenerative reaction is completed and the composite is fabricated. In the sample S₃, the regenerative reactions are not completed even after 48 hours of milling.

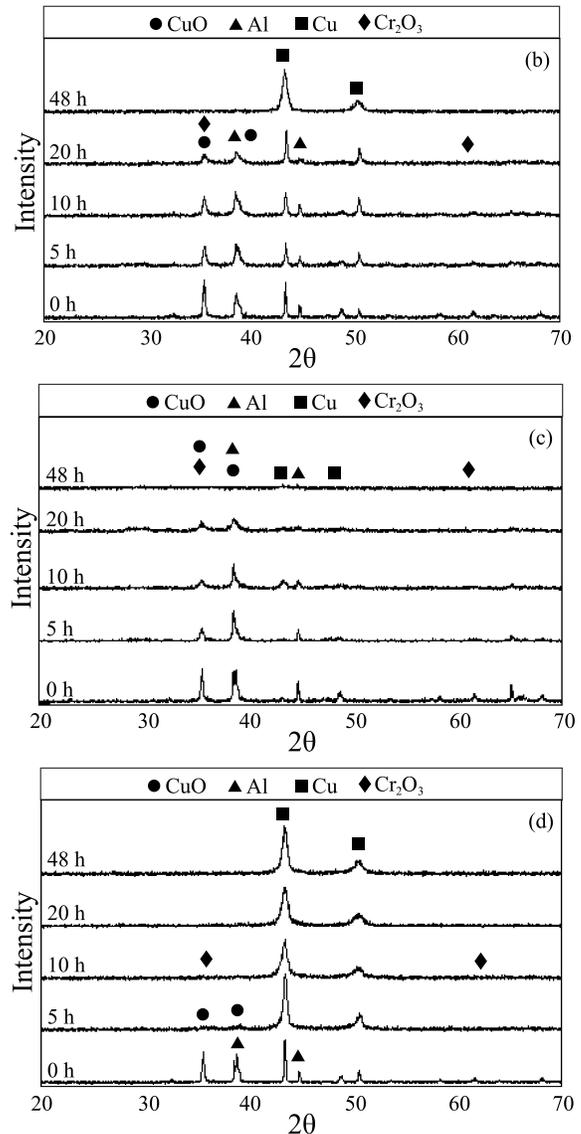
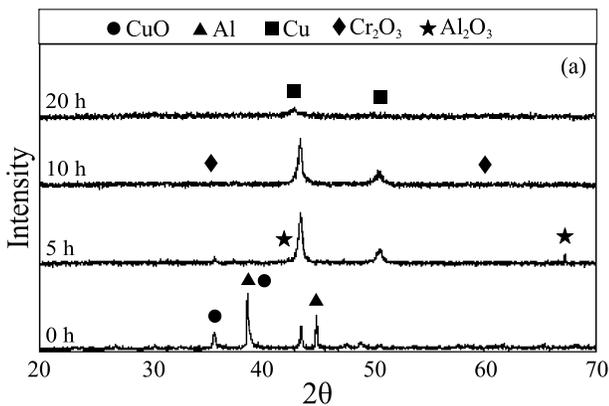


Fig. 1. The XRD pattern of the samples (a) S₁, (b) S₂, (c) S₃, and (d) S₄.

The X-ray diffraction patterns of the sample S₂, which is milled for different periods of time, are illustrated in Fig. 1b. For instance, after five hours of milling, the composition of the phases does not change, and the only considerable phenomenon is that, the peaks of CuO, Al, and Cr₂O₃ shorten due to the mechanical work. As the milling process resumes, the peaks of CuO and Cr₂O₃ gradually weaken, and disappear completely in the 48-hour sample. During the milling process, the peaks of Cu {111} and Cu {200} shorten and widen due to the strain accumulation on the structure and the particle size reduction. Another considerable phenomenon is the movement of copper peaks during the milling process. This commutation resumes approaching sharper angles until 20 hours of milling and returns to the pure copper's position in the 48-hour sample. The commutation phenomenon of the copper peaks can be attributed to the dissolution of the chromium in the copper; since, as can be observed in the equilibrium phase diagram of Cu-Cr, the approximate 8

atomic percent of the chromium in the copper lattice is as a soluble solid solution [10]. The reason of the increase in the copper lattice constant is a distortion which the chromium atoms with the bigger atomic radius in comparison to the copper impose on the unit cell of copper during the substitution dissolution and increase the lattice constant. Therefore, the solution of chromium as a solid solution in the copper lattice can be investigated through measuring the commutation of the copper peaks [11], [12]. Through measuring the distance between the layers d , using the relations between d and the lattice constant (a), using Eqs. (6) and (7) and from the X-ray diffraction pattern, the lattice constant increase can be measured [13]:

$$d = \frac{\lambda}{2 \sin \theta} \quad (6)$$

$$a = d \sqrt{h^2 + k^2 + l^2} \quad (7)$$

From Eqs. (6) and (7), it can be concluded that, the lattice constant increase will lead to the commutation of peak toward sharper angles. With regards to the X-ray diffraction images in Fig. 1, the position of the peak of Cu {111} changes from $2\theta=43.32$ (before milling) to $2\theta=43.12$ after 20 hours of milling. This condition demonstrates that, the lattice constant for the 24-hour-milled sample changes from 3.614 to 3.630. The increase of 0.45% in the lattice constant for the 20-hour-milled sample reflects the fact that, the chromium dissolves in the copper lattice and the substitution solid solution forms (Fig. 2).

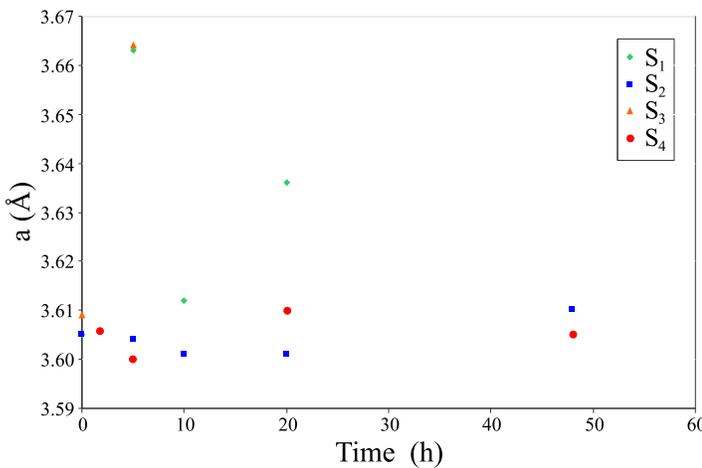


Fig. 2. The changes in the copper lattice's parameter according to the time period of milling.

According to the results of investigating the X-ray diffraction of the milled samples and considering the commutation of the copper peaks and also the disappearance of the peaks of CuO, Cr₂O₃, and Al, it can be concluded that, in the first 24 hours of milling, the chromium dissolves in the copper lattice substitutionally, the solid solution of Cu(Cr) forms and simultaneously, the fracture and crush of oxide particles cause the respective peaks to weaken. As the mechanical alloying

resumes, the size of the particles reduces and the accumulation of structural defects occurs which increases the energy of the system gradually. At this stage (after 20 hours of milling), being provided the thermodynamic conditions, the regenerative substitution reaction starts and the very fine particles of Al₂O₃ form. With the regenerative reaction between the atomic aluminum and the copper, the copper peaks return to the pure copper peaks' position. With regards to the presented mechanism for the reaction, the formed Al₂O₃ will be extremely fine and is not traceable in the X-ray diffraction patterns. The absence of the Al₂O₃ peaks in the final sample is justifiable owing to the same reasons [3], [8].

Comparing the images of Fig. 1, any significant difference cannot be observed in the reaction progress of the samples containing different compositions. In all the three samples with the different percentages of the extra copper, it seems that, the reaction progress is gradual, and even in the sample S₃ with the adiabatic temperature of 2877 K, there is not an abrupt reaction or in the form of MSR which can be recognizable. These results show that, in the prediction of the system's behavior and the determination of the mechanism of the reactions during the milling process, that possesses an intrinsic complexity in which many variables are involved, we cannot simply judge based on the Merzhanov criterion ($T_{ad} > 1800$ K), and this criterion is solely a prerequisite for a pyrogenic reaction to occur as the MSR. This difference in the behavior is more clearly visible when the extra and diluent components are added to the system and the numerous reports about the complexities of the MSR system including a neutral phase support this fact [9]. In this study, the presence of the extra copper in the primary composition, which is added in order to reduce the system's tendency to the combustive reaction, in addition to the confirmation of the predictions of the increasing required time to reach the ignition temperature (T_{ig}), prevents the reaction from performing combustively (the sample S₁). The reasons for this fact can be justified such that; firstly, the high heat capacity of copper increases the heat capacity of the system and decreases T_{ad} and the proportion $\Delta H/C$; secondly, the presence of the copper particles in the primary composition reduces the direct contact between the particles of CuO, Al, and Cr₂O₃, and decreases the reaction rate by positioning between the reacting powders. Overall, the addition of the diluent to the primary materials affects the heat capacity, the collision and contact between the powder particles and the milling container's wall and balls and other parameters. Although using the diluent provides the possibility of controlling the chemical composition in the final product, it causes the reaction to be gradual instead of the MSR.

The changes in the copper grain size (crystal) in the milled samples are obtained from the Cu {111} peaks' width and by means of the Scherer's equation (8) and the William-Hall equation (9):

$$d = \frac{0.9\lambda}{B \cdot \cos \theta} \quad (8)$$

$$B \cdot \cos\theta = \frac{0.9\lambda}{d} + \eta \cdot \sin\theta \quad (9)$$

where d is the crystal size, λ is the X-ray wavelength, B is the peak width at half height, θ is the Bragg peak angle, and η is the lattice strain rate. The remarkable point about the Scherer's equation is that, based on this equation, the widening of the peak is entirely related to the getting tiny of the crystals whereas in the milling process, which the mechanical collisions deeply affect the powder properties, the created strain on the powder particles plays an important role in the widening of the peaks. The considerations demonstrate that, a composition of the two above factors, the getting tiny of the particles and crystals and the particle strain due to the cold working during the milling, has a basic role to play in the widening of the peaks in the X-ray diffraction figures [13]. Now, if $B \cdot \cos\theta$ is plotted according to $\sin\theta$, a straight line will be achieved whose slope equals the lattice strain and its y-intercept will be $0.9\lambda/d$. Therefore, the lattice strain rate can be separated from the changes in the crystal size [15]. The changes in the grain size and the strain rate according to the milling time interval for the sample S_2 are calculated through Eq. (9). The impact of the mechanical alloying on the microstructure and on the crystals' size reduction, also, the effect of the milling process on the deformation phenomenon and on the hardness of work of the particles, which leads to the lattice strain, are illustrated clearly in Figs. 3 and 4. As the milling time interval increases from 5 to 48 hours, the size of the crystals decreases from 120 to 40 nm and the lattice strain rate rises from 0.027 to 0.25. Hence, after 48 hours of milling, the size of the copper crystals becomes nanometric and the composite $Al_2O_3-(Cu-Cr)$ will consequently have a nanometric structure. The SEM images of the powdery mixture with the composition S_2 , before and after the different time periods of milling, are indicated in Fig. 5. The investigation of these images shows that, in the initial time periods of milling, the cold welding occurs and causes the particles to join together and become coarse such that, the size of the agglomerated particles after 10 hours of milling (Fig. 5c) increases within the range of 20-50 μm . As the milling process and the continuous and extensive deformation resume, the hardness of work occurs and results in the fracture and the size reduction of the particles. The considerable size reduction of the particles is this fact's reason. In the 20-hour-milled sample (Figs. 5d and 5e), the size of the particles is less than 20 μm , and in the 48-hour-milled sample, the size of the particles reaches to 30 μm .

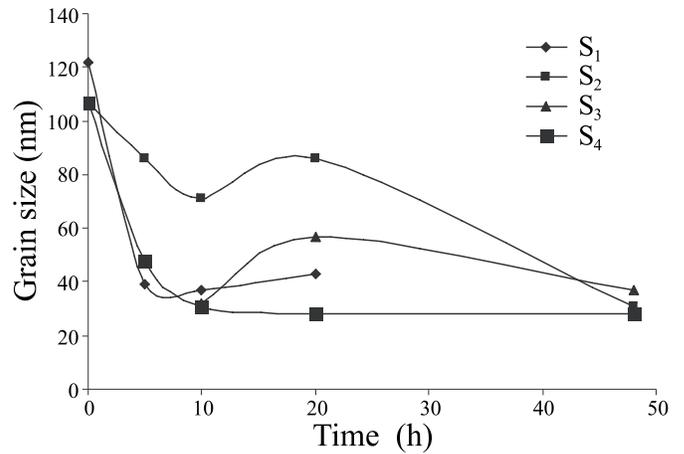


Fig. 3. The change in the crystal size with the time change of milling.

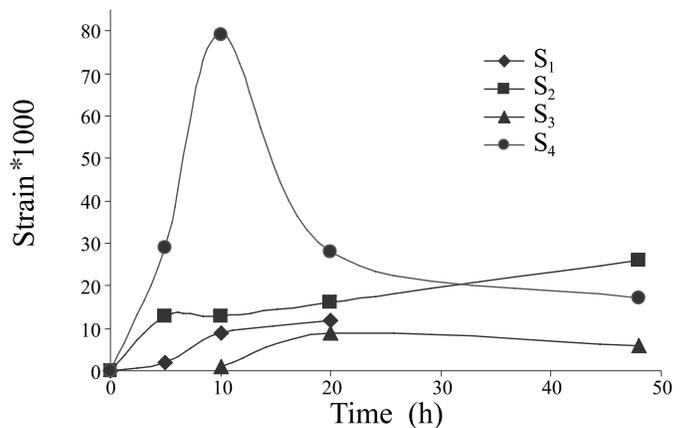
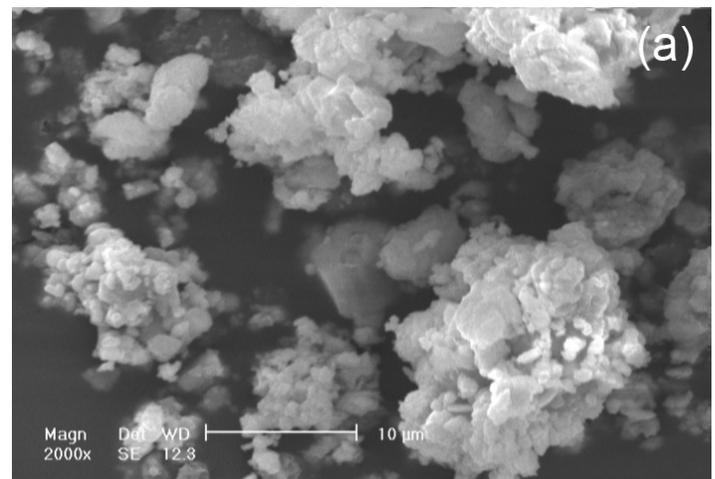


Fig. 4. The change in the lattice strain according to the time change of milling.



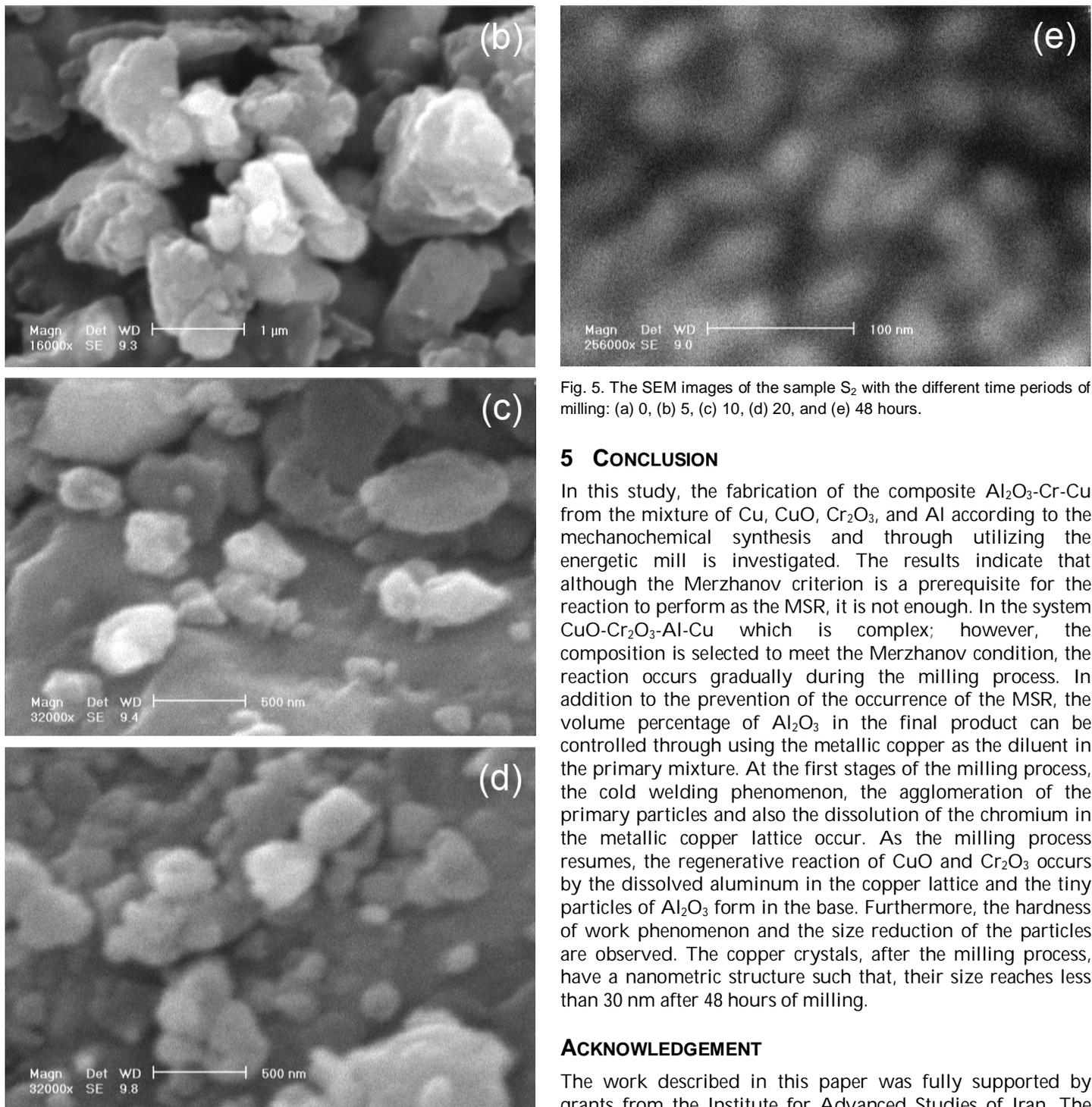


Fig. 5. The SEM images of the sample S_2 with the different time periods of milling: (a) 0, (b) 5, (c) 10, (d) 20, and (e) 48 hours.

5 CONCLUSION

In this study, the fabrication of the composite Al_2O_3 -Cr-Cu from the mixture of Cu, CuO, Cr_2O_3 , and Al according to the mechanochemical synthesis and through utilizing the energetic mill is investigated. The results indicate that although the Merzhanov criterion is a prerequisite for the reaction to perform as the MSR, it is not enough. In the system CuO- Cr_2O_3 -Al-Cu which is complex; however, the composition is selected to meet the Merzhanov condition, the reaction occurs gradually during the milling process. In addition to the prevention of the occurrence of the MSR, the volume percentage of Al_2O_3 in the final product can be controlled through using the metallic copper as the diluent in the primary mixture. At the first stages of the milling process, the cold welding phenomenon, the agglomeration of the primary particles and also the dissolution of the chromium in the metallic copper lattice occur. As the milling process resumes, the regenerative reaction of CuO and Cr_2O_3 occurs by the dissolved aluminum in the copper lattice and the tiny particles of Al_2O_3 form in the base. Furthermore, the hardness of work phenomenon and the size reduction of the particles are observed. The copper crystals, after the milling process, have a nanometric structure such that, their size reaches less than 30 nm after 48 hours of milling.

ACKNOWLEDGEMENT

The work described in this paper was fully supported by grants from the Institute for Advanced Studies of Iran. The authors would like to express genuinely and sincerely thanks and appreciated and their gratitude to Institute for Advanced Studies of Iran.

REFERENCES

- [1] M. S. Motta, P. K. Jena, E. A. Brocchi., and I. G. Solrzano, "Characterization of Cu- Al_2O_3 nano-scale composites synthesized by in situ reduction", Materials Science and Engineering C, Volume 15, Issues 1-2, pp. 175-177, 2001.

- [2] P. K. Jena, E. A. Brocchi, and M. S. Motta, "In-situ formation of Cu-Al₂O₃ nano-scale composites by chemical routes and studies on their microstructures", *Materials Science and Engineering A313*, pp. 180-186, 2001.
- [3] D. Y. Ying, D. L. Zhang, "Solid state reactions between CuO and Cu (Al) or Cu₉Al₄ in mechanically milled composite powders", *Materials Science and Engineering A361*, pp. 321-330, 2003.
- [4] J. R. Gschneidner, L. Eyring, "Handbook on the Physics and Chemistry of Rare Earth", Edited by K. A., Elsevier Science B., Volume 24, 1997.
- [5] D. L. Zhang, J. J. J. Richmond, "Analysis of Compressing and Shearing Behavior of Powders in High-Speed Elliptical-Rotor-Type Powder Mixer (HEM)", *Mater Sci*, Volume 33, pp. 34-70, 1999.
- [6] D. Y. Ying, D. L. Zhang, "Processing of Cu-Al₂O₃ metal matrix nanocomposite materials by using high energy ball milling", *Materials Science and Engineering A*, pp. 152-156, 2000.
- [7] X. Shengqi, Q. Xiaoyan, M. Mingliang, Zhou Jingen, Z. Xiulin, and W. Xiaotian, "Solid-state reaction of Al/CuO couple by high-energy ball milling", *Journal of Alloys and Compounds*, Volume 268, pp. 211-214, 1998.
- [8] S. J. Hwang, J. Lee, "Mechanochemical synthesis of Cu-Al₂O₃ nanocomposites", *Materials Science and Engineering A*, Volume 405, pp. 140-146, 2005.
- [9] L. Takacs, "Self-sustaining reactions induced by ball milling", *Progress in Materials Science*, Volume 405, pp. 140-146, 2005.
- [10] X. J. Liu, I. Ohnuma, R. Kainuma, and K. Ishida, "Phase equilibrium in the Cu-rich portion of the Cu-Al binary system", *Journal of Alloys and Compounds*, Volume 264, pp. 201-208, 1998.
- [11] D. Das, P. P. Chatterjee, I. Manna, and S. K. Pabi, "A measure of enhanced diffusion kinetics in mechanical alloying of Cu-18 at %Al by planetary ball milling", *Scripta Mater*, Volume 41, pp. 861-866, 1999.
- [12] Y. Ying, D. L. Zhang, "Solid-state reactions between Cu and Al during mechanical alloying and heat treatment", *Journal of Alloys and Compounds*, Volume 311, pp. 275-282, 2000.
- [13] B. D. Cullity, "Element of X-Ray Diffraction", Second Edition, Department of Metallurgical Engineering and Materials Science, University of Notre Dame, 1984.
- [14] C. Suryanarayana, "Mechanical alloying and milling", *Progress in Materials Science*, Volume 46, pp. 1-184, 2001.
- [15] B. Fogagnolo, M. Ruiz-Navas, H. Robert, and M. Torralba, "The effects of mechanical alloying on the compressibility of aluminum matrix composite powder", *Materials Science and Engineering A*, Volume 355, pp. 50-55, 2003.

Responsibility on Environmental Damage with a View in the Republic Of Macedonia

Adnan Jashari, PhD

Abstract-The rapid technical and technological development in this century has created many environment harming sources. By hazarding the environment, to a great extent there is an encroachment of human lives, their corporal integrity, their wealth, etc. All of this can result in progeniture of perils for the existence of mankind itself, having in mind that a human being can't develop with a polluted air and water, with waste and hazardous materials all around, etc.

That is why different measures are envisaged, measures whose aim is to interdict or restraint of the environment endangering sources. Since it is a comprehensive global concern, the international community (as well as states in individual order) is undertaking measures in order to preventively affect the swerving of all the dangers and the elimination of the consequences afflicted by the dangers caused previously. In the corpus of undertaken measures are included many international convents whose subject of regulation is the environment. With these Convents above all there is foreseen the possibility of taking responsibility for the damage caused to the environment as well as the criminal-legal responsibility.

Index Terms- civil responsibility, corporate integrity, criminal responsibility, damage, environment responsibility, liability base.

◆

1. THE MEANING OF RESPONSIBILITY (LIABILITY) FOR DAMAGE CAUSED TO ENVIRONMENT

In the environment, the damage can be caused by the activity as well as the negligence of certain people. The damage can be caused by dangerous things or acts. In some cases the damage can be caused also by different people for whom other people take responsibility. The defector may take responsibility for causing the damage, for the dangerous acts or for another person's acts. Depending on the scale of dangerousness and the loftiness of the caused damage, the responsibility for the damage caused may be of civilian and criminal nature. For the criminal responsibility the most important thing is to prove the guiltiness of the defector because without guilt there can't be a penal offence. For the civilian responsibility besides the accountability based on fault, there are cases where the guilt doesn't have to be confirmed as in the case of responsibility for dangerous acts and things. It is the case where the damage is caused by hazardous substances. Hazardous substance is regarded every substance or proprietary that by its traits can be harmful for humans, the environment or property, such are the oxidative and corrosive substances, explosives, etc. The environment can be endangered or damaged from the impact of these substances, hence causing certain detriments.

In cases of damaging the environment, the defector may not be regarded responsible but another person, a person that is in a specific relation to the defector can vouch for him. If an offense has been evoked by the engagement or negligence over the environment then we always bear the individual responsibility of the defector.

2. JURISTIC SOURCES ON RESPONSIBILITY FOR THE DAMAGE CAUSED TO ENVIRONMENT

The prime juristic source of the EU in the scope of responsibility for damaging the environment is the Directive on environmental liability with regard to the prevention and remedying of environmental damage¹. But, even before this Directive was adopted, many issues regarding the environmental liability where included in the so called "White Paper on environmental liability"². The aim of composing such a directive was to ensure preventive measures in regard to environmental damage as well as providing the conditions for remedying the environmental damaging according to the unified rules of environmental liability based on the 'polluter pays' principle. Another important source on this field can be considered the Rio Declaration (1992) which contains many rules, so creating obligations for states to develop their national rights with regards to liability and remedying of environmental damage³.

Besides the Rio Directive there are many more international agreements in the sphere of environment, which as sources of international jurisdiction are imposed to the EU members. Such are the agreements whose objective is the regulation of responsibility for damage

¹ No.2004/35/EC, year 2004.

² White Paper of 9 February 2000 on environmental liability COM (2000) 66 - Not published in the Official Journal]. Commission of the European Communities, Brussels, 9.2.2000, COM (2000) 66 final.

³ For more; P. Sands, Principles of international environmental law, Cambridge University Press, 2003. pp.869-939.

caused by petroleum, dangerous substance transportation as well as the responsibility throughout the utilization of nuclear energy. These kinds of agreements are signed by the Republic of Macedonia with the Republic of Albania for the protection of the Ohrid Lake and the Agreement of cooperation in the field of environment protection with the Federal Republic of Yugoslavia⁴. There are two other Conventions that are of a great importance in the sphere of environmental liability: the Convention on Civil Liability for Damage Resulting from Activities Dangerous to the Environment-1993 and the Convention on the Protection of the Environment through Criminal Law-1998

National sources in Macedonia are: the Constitution, the Law on Protection of the Environment, the Law on Mandatory Relations, the Law on ownership rights in to environmental protection, the Forest Protection Law, the Water Law, the Law on Hunting, the Fishery Law, the Criminal Law, etc. Surely there are many more sub-law acts and other regulation policies which operationalize many issues on the scope of environmental liability.

3. THE TERMS ON DAMAGE CAUSED TO THE ENVIRONMENT

To beget the responsibility on the environmental harm certain conditions ought to exist. Without the fulfillment of certain conditions it cannot be talked about the caused damage. For the environmental liability to arise these conditions are demanded: the caused damage, the capability of liability (offender capability), anti-law action, causal link and the liability base.

3.1. The meaning of environmental damage

The damage is an important term on causing the environmental damage because without damage we can't talk about liability. The damage can induce consequences to properties as well as human life. The damage to property can cause its reduction or block its expansion. The damage on the human individuality can be manifested as in inflicting physical, psychic distress or fear (also known as immaterial damage). So, an environmental damage is considered: the loss of a human life or personal encroachment, the loss or violation of property, the outlay of preventive measures as well as any damage caused through the implementation of these measures.

The environmental damage is defined as damage caused to protected species, water or soil⁵. Damage on protected species is considered if their conservation status

is not fully enforced. Exemptions are indicated, as in the cases where the operator has a delegated permit over the protected species, then the possible negative effects caused to their status won't be classified a damage⁶. Damage caused to water⁷ is considered each harming act caused by different chemical substances. The damage caused to soil implies to the damage caused by various contaminations which can arouse substantial harm to human lives, as a result of direct or indirect effect of different substances on organisms and microorganisms⁸. In general, the damage caused to environment is regarded as a negative alteration of natural resources which directly or indirectly affect the shift of the functions of natural resources. Based on the available data, in Macedonia the damage in most of the cases is caused due to the unsuccessful waste management whose aftereffects reflect on the pollution of water, air as well as the endangerment of human lives. The greatest number of applications for caused damage during the 2009-2011 stretch is made during 2009. While during the year 2009 have been submitted 451 requests for initiating legal process for damage caused to the environment due to poor environment managing, in 2010 are submitted 89 and in 2011 only 26 requests. Analyzed by their content, most of the applications deal with the issue of environment pollution due to bad management of waste. Thus, in a three year period (2009-2011) the total number of requests submitted on account of the mentioned issue reaches approximately 4000 such⁹.

3.2. Criminal liability

For an individual to be held responsible for environmental damage it needs to possess the ability of liability. Liability means legal responsibility for one's acts or omissions. To natural persons it is dependent on age¹⁰. Legal entities have criminal capability so they are legally responsible for the caused environmental damage. They are liable for damage caused by their organs because the organs of a legal entity act on its account.

The principal responsible for the environmental damage are: operators, competent organs as well as the natural persons or legal entities that have been damaged or have a certain interest related to the damage caused to environment.

Operator means the person (natural or legal) who exercises the control of a professional activity¹¹, based on the national provisions, providing that it possesses a permit

⁶ Article 6 and article 16 of the Directive.

⁷ This damage is envisaged in Directive 2000/60/EC.

⁸ D. Wilinon, Environment Law (first edition), New York, 2002, pp. 47

⁹ The data have been secured from the Commission for environmental protection of the Republic of Macedonia.

¹⁰ The Law of obligations, article 160.

¹¹ Professional activities are defined by the Regulation for professional activities, whose practice can induce liability for caused damage "Official Gazette of the Republic of Macedonia" no.31/11.

⁴ Published in the "Official Gazette of the Republic of Macedonia". Nr.46/05. and "Official Gazette of the Republic of Macedonia". nr. 13/03

⁵ J.Mallor, J.Barnes, Th.Bowers, Business law and the Regulatory Environment (tenth edition), Irwin Mc.Graw-Hill, 1995, pp. 1133;

or authorization for such an activity or registered subjects for an activity of that type¹². Liable for the damage caused to the environment is regarded an operator who exercises professional activities and with exercising these activities has caused damage, respectively, potential risk for causing the damage. The operator is liable, if:

- a) Has not undertaken the required preventive measures;
- b) Has not undertaken the remedying measures;
- c) Has not informed the proper organs about the risk of ecological damage that can be caused although the measures have been undertaken¹³.

The cases of the operators that aren't considered liable are defined by specific legal provisions. These are the cases when the operator can testify that the environmental damage is caused by third parties and it has been caused even though all the required security measures have been undertaken. Also, exemptions of operators off the liability are emphasized when the damage is or will be caused in near future as a result of respecting the instructions by the public institutions. The operator is discharged of liability also if the damage is a result of applying the sources (the Directive) and he testifies that he is innocent and that the damage is generated as a result of emitting or various other actions fully permitted by the law. There are such cases in Macedonia also. From a total of 22 cases, 5 cases of discharge are as a result of fully permitted actions made by the defector.

An identical situation (exemption from liability) is applied when the damage is caused by military conflict, civil war as well as circumstances of inevitable natural disasters.

3.3. Anti –legal action

Anti legal action exists when with acts or negligence a juristic regulation is offended. With these acts the material and the non-material interests of certain persons can be offended. Each act or omission which offends the juristic regulations incorporated into the national or international resources is considered an anti-legal act. So, in the Republic of Macedonia, during the three year period (2009-2011) in most of the cases the regulatory provisions of the Law on environment protection, the Law on waste management, the Law on air quality as well as other laws have been offended. From 546 cases of law infringement, most of them are regarding the Law on waste management. In three years (2009-2011) have been found out 451 cases of law infringement regarding the Law on waste management.

3.4. Causal link

The causal link as a stipulation should exist as the liability for the damage caused by their own actions and also the responsibility for the dangerous things or actions. This link exists between acts or omissions and the caused consequence. This link needs to be verified by the endangered. The causal link is an objective category because there doesn't have to be the guilt of the person that is or is not undertaking action.

3.5. Basis of liability

The basis of liability explains the reasons of why a person is responsible for the caused damage. The reason of liability for the caused damage can be the guiltiness or even there are cases when the guilt isn't taken in regard as in the case of causing damage to the environment by dangerous matters or acts. At the liability on environmental damage, the guilt is assumed. This means that if the defector proves that he was not guilty for the caused damage he is not liable. Responsibility without guilt (objective) is emphasized when the environmental damage is caused by harmful acts which at any time can be the reason for the damage on the environment. The industrial and nuclear energy progress are facts which prove that humans and their environment are surrounded by many menaces capable of causing damage to them or their property. The reasonable thing in this case is the one who is initiating and organizing dangerous activities to take the responsibility. These activities are the reason for the environmental damage, where the level of reward depends on the caused damage and not on the fault of the owner, holder or the user of the dangerous matter.

In the Republic of Macedonia, during the period of 2009-2011, the guilt of the defector is considered the base upon which the pronouncement of sanctions is made.

In this direction, the total number of pronounced sentences for damage caused to the environment in this period (three years) is 54614.

4. THE AIM OF LIABILITY FOR ENVIRONMENTAL DAMAGE

The aim of the environmental liability is:

- a) prevention and full remedying of the caused environmental damage,
- b) restitution of the environment, and
- c) providing the measures for minimizing of the risk from environmental damage¹⁵.

Based on the Directive, there are certain types of measures, as: measures dealing with the preventive

¹² See; article 2 of the Directive.

¹³ The law on environmental protection, article 157.

¹⁴ The data have been secured from the Commission for environmental protection of the Republic of Macedonia.

¹⁵ The law on environmental protection, article 157.

activities, measures dealing with environmental damage and the remedying measures¹⁶. The preventive measures are the measures undertaken as a response to the occurrence, activity or incurrance by the procedure and as a result of it a possible risk for causing damage to the environment is created and they are dealing with the preventive or minimization of the damage. The measures of remedying¹⁷ are dealing with measures dedicated to the rehabilitation or replacement of the damaged natural resources.

Within these goals, natural persons or legal entities as well as the civil associations established for the aim of protecting the environment have the legal right to request from the operator in the court:

a) to make the restitution of the environment to its previous state, or

b) to compensate the damage caused to the environment, if the restitution to its previous state is impossible.

If the damage is caused to the public goods, the restitution or compensation can be requested by the Republic of Macedonia. In such case in the name of the Republic of Macedonia the demand to the competent court of law is delivered by the competent national organ for environmental protection.

The carrier of the right can request from the court to order the respondent to secure information or to enable gathering the information from the source of pollution, information needed to determine the responsibility for the ecological damage and the volume of the responsibility¹⁸. Determining the value of the goods and the effect on the environment is made based on the Assessment Methodology¹⁹. Assessing the value of the goods and their effect on the environment is made according to these methods:

a) the Method of assessment based on the market prices, naming the preventive expenses, the expenses of dislocating and repairing the goods, etc.

b) the Method of assessment based on the productive functions, according to which the effects of damage on production, human being, etc, is determined.

c) the Method based on assumed assessment on the caused damage and also the assessment of conclusive expenditure.

No matter which of the methods is used, during the assessment, always the assessment of the goods and the effect on environment is needed, thus, they need to be identified, quantified and assessed correctly. In the Republic of Macedonia, during the period of 2010-2011 have been sentenced 62 cases, with amounts from 30 up to

3000 Euros. Based on the circumstances presented by the respondents, after the scrutiny of the facts 22 persons have been set free of the charge²⁰.

5. REASONS FOR PROTECTION OF THE ENVIRONMENT THROUGH CRIMINAL LAW

Most of the activities by which human beings harm, destroy, pollute or endanger the environment are the result of man's conscious omission or failure to abide by the rules, technical guidelines and standards in managing many dangerous sources of energy and raw materials or in handling hazardous appliances and technologies properly.

Suchlike illegal activities that can be conducted by individuals, groups, trans-national corporations or even entire countries are illegal and punishable as criminal offence and misdemeanors (delicts). With the term "environment offence" we indicate activities and demeanor of people and other subjects that damage the human being and risk general social values. These acts (delicts) are only one of many types of delicts caused in the society. The feature that distinguishes this kind of delict from other types of delicts is the protected object. In this specific case, the protected object is man and one of the fundamental human rights to live in healthy living conditions. This means that the consequences may not be only expressed as physical damage (destruction, smaller or larger magnitude of harm, or activities that make the living and working environment impossible to use). These are the reasons why protection of the environment is also performed by the regulations of criminal law²¹. The intent is that by the pronounced penalties to assist the elimination of conducts that endanger the environment and at the same time enormous revenues are achieved by the perpetrators. These actions are also treated as environmental crimes. In order to secure a quality environment and to secure timely and more effective protection in Macedonia has been created a processed system of environmental misdemeanors. The classifying is made by taking into account their nature, importance and character. Juristic basis for these acts is the Criminal Code. These offenses are classified in two groups: criminal offenses against the environment; and criminal offenses against the environment and public security and property. Nevertheless, the State Assembly has approved some more regulations that ensure environmental protection, as: protection of medical plants, genetically modified organisms. These actions are also included in the Criminal Code. The protected object in all the environmental criminal offences (real, secondary or unreal) is either the environment as a whole or some of its parts:

¹⁶ Article 5 and 7 of the Directive.

¹⁷ These measures are envisaged in the Regulation on remedying measures for ecological damage "Official Gazette of the Republic of Macedonia" no.31/11.

¹⁸ The law on environmental protection, article 159.

¹⁹"Official Gazette of the Republic of Macedonia" no.115/10

²⁰ The data have been secured from the Commission for environmental protection of the Republic of Macedonia.

²¹ These regulations are envisaged in the Directive 99/08/EZ, on Environmental Protection.

air, water or soil, flora and fauna, which are violated or harmed either by the commission of an activity or an omission to act. Yet, the most recent conceptions in the legal theory often point out that the object of protection in these criminal offences is the human right to live in a healthy environment, as one of the fundamental human and universal rights of man and the citizen.

Based on the systematization made to the Criminal Code of Macedonia, it can be noticed that distinctions between different types of environmental criminal offense are made:

1) real environmental criminal offences (i.e. environmental offences in a narrow sense) which are systematized in the Criminal Code of Macedonia within the group of criminal offences against the environment; the immediate object of protection in these offences is the environment as a whole or some of its integral parts that constitute a healthy and natural environment;

2) unreal environmental criminal offences (i.e. environmental offences in a broader sense) which are systematized in the Criminal Code of Macedonia within some other groups of criminal offences; in these offences, the environment comes under attack only indirectly whereas the primary or prevalent object of attack are some other social , and

3) criminal offences envisaged with other legislative acts that regulate the environment as a whole or some of its integral parts.

With the objective of reinforcing the responsibility of offenders (natural persons, legal entities, trans-national corporations or the state), in the scope of environment protection in international level is adopted the Convent which is considered the foundation of criminal liability for environmental damage. In the Republic of Macedonia there is not initiated any process of criminal nature.

6. THE PERSPECTIVE OF ENVIRONMENTAL PROTECTION IN THE REPUBLIC OF MACEDONIA

Given the importance of environment to the development of society and other living beings, especially when knowing that there is a continuous tendency of environment pollution, as a prominent issue is raised the necessity of generating mechanisms for environmental protection with higher efficiency, moreover, for a more rapid and efficient compensation for damage caused to the environment. All this is determined by the fact that nowadays the number of people that are carrying out dangerous actions whose perpetration can at any time cause damage to the environment is in rise. On the other hand, knowing that the procedures of compensation are evolved to a certain degree so there always exists the

possibility of attacking the primary verdicts with appeal as well as utilizing extraordinary legal means for whose application time and expenses are inalienable.

We are assessing the importance of mechanisms in the scope of environmental protection as:

a) firstly, ensuring the liability, is a very good and effective mechanism in protecting the environment. This ensuring must have an obligatory nature, and in this regard any operator should not be handed a permit if he/she previously has not contracted an insurance covenant with a certain insurance company. Provided the assurance of environmental liability regulations the operator is also protected from possible financial consequence in case of causing damage and at the same time the process of damage compensation is fulfilled efficiently and avoiding financial expenditure and long procedures;

b) secondly, the scale of dangerousness defined by juristic provisions in Macedonia for a environmental delict to be rated a criminal offence needs to be as low as possible. This is indicated through the current situations in Macedonia where from all the acts and omissions that have caused damage to the environment; none of them has been regarded as a criminal offence. This fact shows that damagers of the environment nevertheless fear the prescribed sanctions.

CONCLUSION

The process of industrialization has caused many positive as well as negative consequences to mankind. As a negative consequence is regarded the possibility of continuous increase in pollution of environment caused by industrial development.

Therefore the pollution of environment is a very serious problem facing society nowadays. All of this has influenced the international as well as the national institutions to take measures aiming to detain or reduce the impact of industries on environment. Among the most important measures provided for the protection of environment are regarded the juristic measures adopted by the international and national institutions. Based on these legal acts (national and international) a good juristic range on environmental protection in the criminal-juristic and in the civil-juristic scheme as well as in other spheres with administrative-juristic nature has been created.

REFERENCES

- [1] D.Wilinson, Environment Law (first edition), New York, 2002.
- [2] J.Mallor, J.Barnes, Th.Bowers, Business Law and the Regulatory Environment (tenth edition), Irwin Mc.Graw-Hill, 1995.

- [3] P. Sands, Principles of international environmental law, Cambridge University Press, 2003.
- [4] P. Stookes, A practical approach to environment law, Oxford University Press, 2009.
- [5] P.Sulejic, Pravo osiguranja, Beograd, 2005.
- [6] G.Galev, Obligaciono pravo, Skopje, 2009.
- [7] Regulation on remedying measures for ecological damage "Official Gazette of the Republic of Macedonia" no.31/11.
- [8] Regulation for professional activities, whose practice can induce liability for caused damage "Official Gazette of the Republic of Macedonia" no.31/11
- [9] White Paper of 9 February 2000 on environmental liability COM (2000) 66 - Not published in the Official Journal]. Commission of the European Communities, Brussels, 9.2.2000, COM (2000) 66 final:
- [10] White Paper on environmental liability, Nr.2004/35/EC, year 2004:
- [11] Directive regarding preventive and remedying for caused environmental damage, No.2004/35/EC, year 2004:
- [12] Convention on Civil Liability for Damage Resulting from Activities Dangerous to the Environment (1993):
- [13] Convention on the Protection of the Environment through Criminal Law (1998):
- [14] Law on environmental protection "Official Gazette of the Republic of Macedonia", no.51/11
- [15] Law on waste management "Official Gazette of the Republic of Macedonia", no.9/11.
- [16] Law on air quality "Official Gazette of the Republic of Macedonia", no. 47/11.
- [17] Law of obligations, "Official Gazette of the Republic of Macedonia", no.17/05.
- [18] Information about protection of the environment in the Republic of Macedonia, years: 2008, 209, 2010, 2011.
- [19] Report of the Commission for environmental protection of the Republic of Macedonia, years: 2009, 2010, 2011.

Rare Association Rule Mining using Improved FP-Growth algorithm

T. Ravi Kumar¹

¹Associate Professor, Department of CSE
Swarna Bharathi Institute of Science & Technology, Khammam,
¹trkumar5@gmail.com

K. Raghava Rao²

²Professor in CSE, School of Computing, KL University, Vaddeswaram, Guntur, A.P.
INDIA¹²
²raghavarao1@yahoo.com

Abstract: Rare association rule refers to an association rule forming between frequent and rare items or among rare items. CFP-growth approach is used to mine frequent patterns using multiple minimum support (minsup) values. This approach is an extension of FP-growth approach to multiple minsup values. This approach involves construction of MIS-tree and generating frequent patterns from the MIS-tree. The issue in CFP-growth is constructing the compact MIS-tree because CFP-growth considers certain items, which will generate neither frequent patterns nor rules. In this paper, we propose an efficient approach for constructing the compact MIS-tree. To do so, the proposed approach explores the notions “least minimum support” and “infrequent child node pruning. The proposed approach improves the performance over CFP-growth approach.

Keywords: Data mining, Association rules, IMSApriori, CFP-growth and ICFP-Growth

1.Introduction

Data mining represents techniques for extracting patterns hidden in large datasets by combining methods from statistics and artificial intelligence with database management. In Data mining approaches mainly focuses on discovering the frequent occurring items and the knowledge pertaining to it. However, real-world datasets are mostly non-uniform in nature containing both frequent and relatively infrequent or rarely occurring entities. But rare knowledge patterns are more difficult to detect because they present in fewer data cases.

In this paper, we are proposing an improved approach to extract rare association rules. Rare association rule refers to an association rule forming between frequent and rare items or among rare items. Association rule generation is usually split up into two separate steps:

1. First, minimum support is applied to find all frequent item sets in a database.
2. Second, these frequent item sets and the minimum confidence constraint are used to form rules.

Minsup controls the minimum number of data cases that each set of items must cover. Minconf controls the predictive strength of a rule. The main issue in mining rare association rules is finding the frequent patterns involving both frequent and rare items. And the later are not found in the algorithms like Apriori and FP-Growth as they suffer with "rare item problem". That is, at high minsup value, frequent patterns involving rare items could not be extracted as rare items fail to satisfy the minsup value. Therefore in order to find the rare items the minsup value would be set to low which also results in producing many frequent items.

1.1System model and notations

The problem of association rule mining would be defined as: Let $I = \{i_1, i_2, \dots, i_n\}$ be a set of n binary attributes called items. Let $D = \{t_1, t_2, \dots, t_m\}$ be a set of transactions called the database. Each transaction in D has a unique transaction ID and contains a subset of the items in I. A rule is defined as an implication of the form $X \Rightarrow Y$ where $X, Y \subseteq I$ and $X \cap Y = \emptyset$. The sets of items X and Y are called antecedent and consequent of the rule respectively. The rule $X \Rightarrow Y$ holds in D with support s, if s% of the transactions in D contain X \Rightarrow Y. Similarly rule $X \Rightarrow Y$ holds in D with confidence c, if c% of transactions in D that support A also support B.

2.MULTIPLE MINIMUM SUPPORT APPROACHES

2.1 Minimum item support

In multiple minimum support based frequent pattern mining, each item is specified with a minsup value called minimum item support (MIS). Frequent items are the items having support greater than or equal to their respective MIS values. Infrequent items are the items having support less than their respective MIS values.

$$S(i_1, i_2, \dots, i_k) \geq \min \left(\begin{matrix} MIS(i_1), MIS(i_2) \\ \dots, MIS(i_k) \end{matrix} \right)$$

MIS for each item is calculated as follows

$$MIS(i_j) = \beta S(i_j), \text{ if } \beta S(i_j) > LS \\ = LS \text{ else}$$

Where SD refers to support difference

$$SD = ! (1 \#)$$

Where ! represents the parameter like mean, median, mode, maximum support of the item supports and # is the parameter ranging between 0 to 1. SD takes values from (0, !).

2.2 CFP-Growth approach

In order to address the performance problems involved in MSApriori approach, an efficient approach known as CFP-growth. The CFP-growth is an extension of single minsup based FP-growth approach to multiple minsup values. This approach involves two steps. They are construction of MIS-tree and mining frequent patterns from the MIS-tree using conditional pattern bases. This approach assumes that information regarding the MIS values for the items will be the provided by the user priori to its execution., the MIS-tree is constructed as follows.

1. First, the items are sorted in descending order of their MIS values, say L1 and their frequency values are set at zero. Next, a root node of the tree is constructed by labeling with "null". Next, for each transaction in the dataset the following steps are performed to generate MIS-tree. They are:

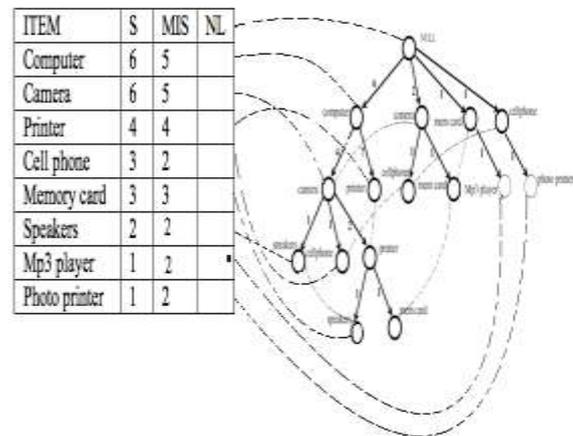
2. A branch is created for each transaction such that nodes represent the items, level of nodes in a branch is based on the sorted order and the count

TID	Items
1	Computer, camera, speakers
2	Camera, cell phone
3	Camera, memory card, printer
4	Computer, camera, cell phone
5	Computer, printer
6	Camera, memory card
7	Computer, printer
8	Computer, camera, printer, speakers
9	Computer, printer, camera, memory card
10	Memory card, mp3 player
11	Cell phone, photo printer

of each node is set to 1. However, in constructing the new branch for a transaction, the count of 1. The items in each transaction are sorted in L1 order. Next, update the frequencies of the items which are present in the transaction by incrementing the frequency value of the respective item by 1. each node along a common prefix is incremented by 1, and nodes for the items following the prefix are created,

linked accordingly and their values are set to 1.

To facilitate tree traversal, an item header table is built so that each item points to its occurrences in the tree via a chain of node-links. From the item frequencies, the respective support values are calculated. Using the lowest MIS value among all the items (MIS), the tree-pruning process is performed on the item header and MIS-tree to remove the items having support is less than the lowest MIS value. After tree pruning, tree-merging process is



Mining the constructed MIS-tree is shown in Table2.

In multiple minsup based frequent pattern mining infrequent items can still generate frequent patterns. For a pattern to be frequent, it should satisfy the lowest MIS value among the items in a pattern. So, given the MIS values of every item, any frequent pattern will have support greater than or equivalent to the lowest MIS value among all the items. Therefore, if an infrequent item has support less than the lowest MIS value, it will not generate any frequent pattern and hence can be pruned. This notion is exploited by CFP-growth approach to construct the efficient MIS-tree. We illustrate this scenario in Example 1

Table 2: Mining the MIS-tree by creating conditional pattern bases in CFP-growth.

Item	MIS	Conditional pattern base	Conditional MIS-tree	Frequent patterns generated
Mem card	3	{{camera:1,printer:1,computer:1}, {camera:1}}	<camera: 2>	{camera, mem card: 2}
Cell phone	2	{{camera:1,computer:1}, {camera}}	<camera: 2>	{Cell phone, camera:2}
Printer	4	{{Computer:2},{printer:1}, {computer:2,camera:2}}	<computer: 4>	{Printer, computer: 4}
Camera	5	{{Computer:4}}	<computer: 4>	{Computer, camera: 4}

Example1: Let the set of items {i,j,k,l,m,n,o,p} have support values {10%,8%,6%, 6%,4%,3%, 2%, 1%} respectively. Let the respective MIS values be {9%, 9%, 7%,7%, 4%, 3%, 3%,2%}. Since the lowest MIS value is 2%, any frequent pattern will have support not less than 2%. So based on downward closure property, the item 'p' will not generate frequent pattern and henceforth can be removed. Therefore, CFP-growth considers set of items {i,j,k,l,m,n,o,p} for generating frequent patterns. However, it can be observed that the infrequent item 'o' will never generate any frequent pattern. The reason is that any pattern involving item 'o' can have support at most equivalent to 2%, which is less than the MIS value of 'o' i.e., 3% constructed by CFP-growth approach.

3 PROPOSED APPROACH

The proposed algorithm generalizes the CFP-growth algorithm for finding frequent patterns. We call the new algorithm as Improved Complete Frequent Pattern-growth (ICFP-growth). The ICFP-growth approach involves two steps. They are constructing the MIS-tree and generating frequent patterns from the MIS-tree. The ICFP-growth explorer the notions "least minimum support" and "infrequent child node pruning" for constructing the MIS-tree, so that the size of the resultant MIS-tree may be less than or equivalent to the MIS-tree constructed by CFP-growth approach.

3.1 Least Minimum Support

The frequent patterns mined using multiple minsup values follow "sorted closure property" says, in a frequent pattern, all the supersets involving the item having lowest MIS value should be frequent. So in every frequent pattern, frequent item represents the item having the lowest MIS value. Therefore, every frequent pattern will have support greater than or equal to lowest MIS value among all the frequent items. Thus, if we remove all the items whose support is less than the lowest MIS value of the frequent item, no frequent pattern will be missed. This notion is called "least minimum support" (LMS) refers to the lowest MIS value among all the frequent items.

The significance of this notion is illustrated in example3

Example2: Continuing with the Example 1, it can be observed that the set of items {i,m,n} are frequent items. The lowest MIS value among these items is 4%. Therefore, using LMS value as 4%, the proposed approach prunes the set of items {o,p} and considers {i,j,k,l,m} for frequent pattern mining. Let I be the set of all items in the transaction dataset. Let C be the set of items considered by CFP-growth approach for mining frequent patterns. Let F be the set of items considered by ICFP-growth approach for mining frequent patterns. Then, the relation between I, C and F is as follows: $F \subseteq C \subseteq I$.

3.2 The algorithm

The ICFP-growth pre-assumes that for every item, user specifies the MIS values priori to its execution. Therefore, using the priori information i.e., MIS values of the items, the frequent patterns are generated with a single scan on the dataset.

3.2.1 Constructing MIS-tree

The construction of MIS-tree in ICFP-growth algorithm is shown in Algorithm 1 and described as follows. The ICFP-growth algorithm accepts transaction dataset (Trans), Item set (I) and minimum item support values (MIS) of the items as input parameters. Using the input parameters, the ICFP-growth creates an initial MIS-tree, which is similar to MIS-tree created by CFP-growth (Lines 1 to 7 in procedure 1). Next, starting from the last item in the item-header table (i.e., item having lowest MIS value) perform tree-pruning by calling MisPruning procedure (See, procedure 3) to remove the infrequent items from the item-header table and MIS-tree. After one item is pruned, then move to next item in item-header table and perform tree pruning. However, stop tree-pruning process when the frequent item is encountered. The MIS value of this frequent item is the LMS value. Let the resultant item header table be MinFrequentItemHeaderTable. Call MisMerge procedure (See, procedure 4) to merge the tree. Finally, call InfrequentChildNodePruning procedure (See, procedure 5) to prune the infrequent child nodes in the MIS-tree. The resultant MIS-tree is the compact MIS-tree.

3.2.2 Mining frequent patterns from MIS-tree

Mining the frequent patterns from the compact MIS-tree is shown in Algorithm 6. The process of mining the MIS-tree in ICFP-growth is almost same as mining the MIS-tree in CFP-

growth. However, the variant between the two approaches is that before generating conditional pattern base and conditional MIS-tree for every item in the header of the Tree, the ICFP-growth approach verifies whether the suffix item in the header of the Tree is a frequent item (Line 2 in Procedure 6). If an suffix item is not a frequent pattern then the construction of conditional pattern base and conditional MIS-tree are skipped. The reason is as follows. In every frequent pattern, the item having lowest MIS value should be a frequent item. In constructing the conditional pattern base for a suffix item, the suffix item represents the item having lowest MIS value. Therefore, if the suffix item is an infrequent item then the patterns in which it represents the item having lowest MIS value will also be infrequent. So, for an infrequent suffix item, it is not necessary to construct conditional pattern base.

3.2.3 Infrequent child node pruning

The ICFP-growth approach skips the construction of conditional pattern bases for the infrequent suffix items. So in the compact MIS-tree, the child nodes belong to infrequent items have no significance because its prefix paths (conditional pattern bases) are not used. So, if we can prune the child nodes belonging to infrequent items, the resultant MIS-tree will still preserve the transaction details pertaining to frequent patterns. So, in the MIS-tree, “infrequent child node pruning” is performed such that every branch ends with the node of a frequent item. pruning should be performed only on the child nodes belonging to infrequent items. We illustrate the “infrequent child node pruning” in Example 3.

Procedure 1: MIS-tree(Tran: transaction dataset, I: item set containing n items, MIS: minimum item support values for n items)

```

-----
1: Let L represent the set of items
   sorted in non-decreasing order of their
   MIS values.
2: create the root of a MIS-tree, T, and
   label it as "null".
3: for each transaction t # Tran do
4:   sort all the items in t in L order.
5:   count the support values of any
   item i, denoted as S(i) in t.
6:   Let the sorted items in t be [p |P],
   where p is the first element and P
   is the remaining list.
   Call InsertTree([p |P], T).
7: end for
8: for j=n-1; j $ 0; -j do
9:   if S[j] < MIS[j] then
10:    Delete the item ij in header table.
11:    call MisPruning (Tree, L[j]).
12:   else
13:    break; //come out of pruning step.
14:   end if
15: end for
16: Name the resulting table as
   MinFrequentItem-HeaderTable.
17: Call MisMerge(Tree).
18: Call
   InfrequentChildNodePruning(Tree).

```

Procedure 2 InsertTree ([p|P, T)

```

-----
1: while P is nonempty do
2:   if T has a child N such that
   p.item-name=N.item-name then
3:     N.count++.
4:   else
5:     create a new node N, and let its count be 1;
6:     let its parent link be linked to T.
7:     let its node-link be linked to the
   nodes with the same item-name
   via the node-link structure;
8:   end if
9: end while

```

Procedure 3 MisPruning (Tree, ij)

```

-----
1: for each node in the node-link of ij in Tree do
2:   if the node is a leaf then
3:     remove the node directly;
4:   else
5:     remove the node and then its parent
   node will be linked to its child node(s);
6:   end if
7: end for

```

Procedure 4 MisMerge (Tree)

```

-----
1: for each item ij in the
   MinFrequentItemHeaderTable do
2:   if there are child nodes with the
   same item name then
3:     merge these nodes and set the count as
   the summation of these nodes counts.
4:   end if
5: end for

```

Procedure 5

InfrequentChildNodePruning (Tree)

```

-----
1: choose the last but one item ij in
   MinFrequentItemHeaderTable. That is,
   item having second lowest MIS value.
2: repeat
3:   if ij item is infrequent item then
4:     using node-links parse the
   branches of the Tree.
5:   repeat
6:     if ij node is the child of a branch then
7:       drop the node-link connecting
   through the child branch.
8:     create a new node-link from the
   node in the previous branch to node
   in the coming branch.
9:     drop the child node in the branch.
10:    end if
11:   until all the branches in the tree are
   parsed
12:   end if
13: choose item ij which is next in the
   order.
14: until all items in
   MinFrequentItemHeaderTable are
   completed

```

Procedure 6 ICFP-growth (Tree: MIS-tree, L: set of quasi-frequent items, MIS: minimum item support values for the items in L)

```

-----
1: for each item ij in the header of the
   Tree do
2:   if ij is a frequent item then
3:     generate pattern %=ij!# with
       support = ij.support;
4:   construct %'s conditional pattern base
       and %'s conditional MIS-tree Tree %
5:   if Tree % $0 then
6:     call CpGrowth(Tree %, %, MIS(ij)).
7:   end if
8: end if
9: end for

```

Procedure7 CpGrowth(Tree, #, MIS(#))

```

-----
1: for each ij in the header of Tree do
2:   generate pattern %=ij!# with
       support = ij .support;
3:   construct %'s conditional pattern base
       and %'s conditional MIS-tree Tree%
4:   if Tree% $0 then
5:     call CpGrowth(Tree%,%,MIS(#)).
6:   end if
7: end for

```

Example3: Continuing with example 1,2, let the MIS-tree derived after performing a single scan on the dataset contain three branches, say !i,j", !j,k" ,!k,l",!l,m",!m,n". Among the set of items {i,j,k,l,m,n}, we know that items j,k,l are infrequent items. First, let us consider the item l, having relatively lowest MIS value for pruning..This process is repeated until all infrequent items are pruned. Thus the final resulted MIS-tree contains only two branches !i",!l,m" and !m,n".

4. Conclusion

In this paper, we have proposed an efficient algorithm by using the notions "least minimum support" and "infrequent child node pruning" so that the size of the resultant MIS- tree may be less than or equivalent to the MIS-tree constructed by CFP-growth approach. ICFP-growth approach overcoming the performance problems in IMSApriori.The memory requirements of ICFP-growth approach will never exceed those of CFP-growth approach. Mostly, it requires relatively less memory.

References

- 1.Agrawal, R. and Srikanth, R. (1994). *Fast algorithms for mining association rules*. In VLDB.
- 2.B. Liu, W. H. and Ma, Y. (1999).*Mining association rules with multiple minimum supports*. In SIGKDD Explorations.
- 3.ErayOzkural, C. A. (2004). *A space optimization for fp-growth*. InFIMI04.
- 4.G. Melli, R. Z. O. and Kitts,B.(2006).*Introduction to the special issues on successful real-world data miningapplications*. InSIGKDDExplorations, volume8
- 5.H. Jiawei, P. Jian, Y. Y. andRunying, M. (2004).*Mining frequent patterns without candidate generation: A frequent-pattern tree approach*.InDMKD.
- 6.Kiran,R.U.andReddy,P.K.(2009).*An improved multiple minimum support based approach to mine rare association rules*.
- 7.In IEEE Symposium on Computational Intelligence and DataMining. Mannila,H.(1997).
- 8.R. Agrawal, T. I. andSwami, A. (1993). *Mining association rules between sets of items in large databases*. In SIGMOD.
- 9.Ya-HanHu, Y.-L. C. (2004). *Mining association rules with multiple minimum supports: A new algorithm and a support tuning mechanism*. In Decision Support Systems.
- 10.Kiran, KrishnaReddy,(2010), *Rare association rule mining using FP-growth algorithm*.



T Ravi Kumar received M.Sc.(Comp.Sc.) from Andhra University, and M.Tech(CSE) from JNTU Hyderabad. He has got 15 years experience in teaching for UG and PG students in several colleges. His area of research interests are Data Mining, Wireless sensor networks and Web Technologies. Currently he is doing research in Web data mining.



K. Raghava Rao, Professor in CSE, having 12 years experience in teaching for UG and PG students and 3 years of Software Development experience in Singapore. He received B.E(CSE) from MU Univeisty, M.Tech(CSE) from RVP Univeristy, Udaipur and Ph.D(CSE) from MahatamaGandhiUniveristy(Kasividyapeeth), Varanasi, in the years 1995, 2005 and 2009 respectively. He published serveral papers in national & international conferences and journals. His research inetests are Data Mining, Wireless sensor networks and Web Technologies.

ECG Signal Compression Techniques for Data Storage and Qualitative Analysis

Nikhil Dhawan, Saurabh Gupta, Ravi Sadrani, Swati Sanghvi

Abstract-ECG (electrocardiogram) is a test that measures the electrical activity of the heart. The information obtained from an electrocardiogram can be used to discover different types of heart disease. It may be useful to see how well the patient is responding to treatment. An ECG trace is a digitized version of a continuous signal. To reduce the loss in ECG signal we have used some efficient techniques in our project. Various techniques can be used for compression like the Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) etc. ECG signal being used in a wide variety of biomedical applications requires accurate results, less power requirements, faster results and low cost maintenance. Therefore compression plays a very important role in acquiring these purposes without losing the original information. In general, most of the introduced ECG compression techniques have inaccuracy and random behavior of error. Hence a new technique was proposed called as Discrete Wavelet Transform (DWT). Also from the results and computations that we have performed in our project we come to a conclusion that DWT is a better compression technique than DCT since it has better accuracy and also it correlates very well with the subjective tests.

Index terms-Accuracy, compression, Discrete Cosine Transform (DCT), Discrete Wavelet Transforms (DWT), fast, percent root mean square difference (PRD) and without losing original information

1 INTRODUCTION

Electrocardiography (ECG or EKG) is a transthoracic interpretation of the electrical activity of the heart over time captured and externally recorded by skin electrodes. It is a noninvasive recording produced by an electrocardiographic device. The etymology of the word is derived from the Greek electro, because it is related to

Electrical activity, kardio, Greek for heart, and graph, a Greek root meaning "to write". Electrical impulses in the heart originate in the sinoatrial node and travel through the intrinsic conducting system to the heart muscle. The impulses stimulate the myocardial muscle fibers to contract and thus induce systole. The electrical waves can be measured at electrodes placed at specific points on the skin. Electrodes on different sides of the heart measure the activity of different parts of the heart muscle [1],[2],[3]. An ECG displays the voltage between pairs of these electrodes, and the muscle activity that they measure, from different directions, can also be understood as vectors. This display indicates the overall rhythm of the heart and weaknesses in different parts of the heart muscle. It is the best way to measure and diagnose abnormal rhythms of the heart, particularly abnormal rhythms caused by damage to the conductive tissue that carries electrical signals, or abnormal rhythms caused by electrolyte imbalances.^[3] In a myocardial infarction (MI), the ECG can identify if the heart muscle has been damaged in specific areas, though not all areas of the heart are covered. The ECG cannot reliably measure the pumping ability of the heart, for which ultrasound-based (echocardiography) or nuclear medicine tests are used. Compression is the reduction in size

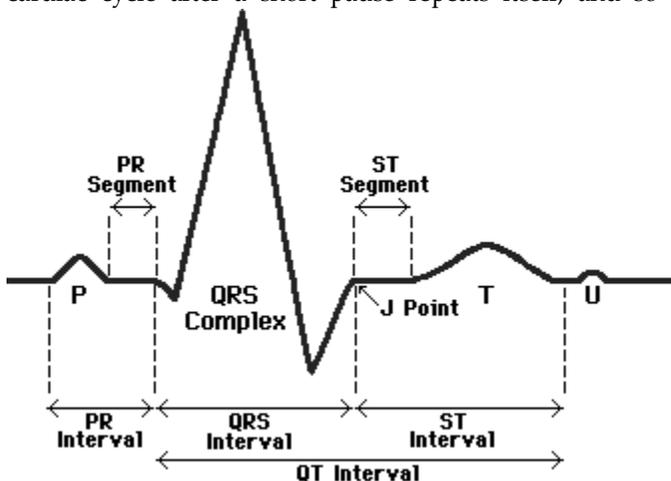
of data in order to save space or transmission time. For data transmission, compression can be performed on just the data content or on the entire transmission unit (including header data) depending on a number of factors. Compression is useful because it helps reduce the consumption of expensive resources, such as hard disk space or transmission bandwidth. Data compression methods for ECG signals have been playing an important role in computer processing and analysis of ECG. The major reason for going in for a higher compression ratio (CR) has been the desire to obtain higher density storage in medical database and hospital information systems [7].

1.1 Diagnosis based on heart signals

Normal sinus rhythm:

In a normal heart rhythm, the sinus node generates an electrical impulse which travels through the right and left atrial muscles producing electrical changes which is represented on the electrocardiogram (ECG) by the P-wave [1],[2],[3]. The electrical impulses then continue to travel through specialized tissue known as the atrioventricular node, which conducts electricity at a slower pace. This will create a pause (PR interval) before the ventricles are stimulated. This pause is helpful since it allows blood to be emptied into the ventricles from the atria prior to ventricular contraction to propel blood out into the body. The ventricular contraction is represented electrically on the ECG by the QRS complex of waves. This is followed by the T-wave which represents the electrical changes in the ventricles as they are relaxing. The

cardiac cycle after a short pause repeats itself, and so on.



Therefore, on an ECG in normal sinus rhythm p-waves are followed after a brief pause by a QRS complex, then a T-wave.



Normal sinus rhythm not only indicate that the rhythm is normally generated by the sinus node and traveling in a normal fashion in the heart, but also that the heart rate, i.e. the rate at which the sinus node is generating impulses is within normal limits. There is no one normal heart rate, but this varies by age. It is normal for a newborn to have a heart rate up to 150 beats per minute, while a child of five years of age may have a heart rate of 100 beats per minute. The adult's heart rate is even slower at about 60-80 beats per minute[1],[2],[3],[6].

Sinus tachycardia:

A fast heart rate may occur with a normal heart rhythm, this is called sinus tachycardia[6]. This means that the impulse generating the heart beats are normal, but they are occurring at a faster pace than normal. This is termed sinus tachycardia and is seen normally with exercise or with excitement.

Sinus bradycardia:

The heart may slow down, yet maintain the normal pattern of rhythm (sinus), this is known as sinus bradycardia [6]. It usually is benign and may be caused by medications such as beta blockers.

Premature atrial contraction (PAC):

The atria fire an early impulse which causes the heart to beat earlier causing irregularity in the heart rhythm.

Premature ventricular contraction (PVC):

The ventricles fire an early impulse which causes the heart to beat earlier causing irregularity in the heart rhythm.

2 DIFFERENT COMPRESSION TECHNIQUES&QUALITATIVE ANALYSIS

Data compression methods for ECG signals have been playing an important role in computer processing and analysis of ECG. The major reason for going in for a higher compression ratio (CR) has been the desire to obtain higher density storage in medical database and hospital information systems. Another area wherein the need for efficient data compression for ECG has been felt is Ambulatory ECG Monitoring (AECGM). AECGM is usually done using the conventional Holter monitor which consists of a 24 hour cassette recording of the ECG. Modern Holter monitors with digital IC memory cards are now expected to improve fidelity of recording and make the system more compact [1],[2],[3],[6]. But due to the limited capacity of the IC memory cards, the sampled ECG data generated during the 24 hours has to be first compressed before it can be stored digitally. Another application which has been proposed recently is compression of ECG data for storing (along with other patient data) on a medical smart card. These and many other applications demand data compression algorithms with very high CRs.

Existing ECG data compression techniques have been classified into

a) Direct data handling:

The direct data handling techniques achieve data compression by removing the redundancies present in the actual ECG signal samples. Techniques such as AZTEC, CORTES, SAPA and DPCM and entropy coding come under this category.

b) Transformation techniques:

In contrast the transform domain techniques achieve data compression by constraining their basis functions. Many discrete orthogonal transforms such as KLT, DCT, and FT have been used for ECG data compression.

2.1 Discrete Cosine Transform

The discrete cosine transform (DCT), closely related to the DFT. The DCT's energy compaction properties are useful for applications like signal coding. The toolbox function DCT computes the unitary discrete cosine transform, or DCT, for an input vector or matrix. Mathematically, the unitary DCT of an input sequence x is

$$y(k) = w(k) \sum_{n=1}^N x(n) \cos \frac{\pi(2n-1)(k-1)}{2N} \quad (1)$$

$$k = 1, 2 \dots N$$

$$\text{Where } w(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 1 \\ \sqrt{\frac{2}{N}}, & 2 \leq k \leq N \end{cases}$$

The DCT is closely related to the discrete Fourier transform; the DFT is actually one step in the computation of the DCT for a sequence. The DCT, however, has better energy compaction properties, with just a few of the transform coefficients representing the majority of the energy in the sequence. The energy compaction properties of the DCT make it useful in applications such as data communications [5],[7].

The function IDCT computes the inverse DCT for an input sequence, reconstructing a signal from a complete or partial set of DCT coefficients. The inverse discrete cosine transform is

$$x(n) = \sum_{k=1}^N y(k) \cos \frac{\pi(2n-1)(k-1)}{2N} \quad (2)$$

$$n = 1, 2 \dots N$$

where

$$\omega(n) = \begin{cases} \frac{1}{\sqrt{N}}, & n = 1 \\ \sqrt{\frac{2}{N}}, & 2 \leq n \leq N \end{cases}$$

Because of the energy compaction mentioned above, it is possible to reconstruct a signal from only a fraction of its DCT coefficients. For example, generate a 25 Hz sinusoidal sequence, sampled at 1000 Hz.

2.2 Discrete Wavelet Transform

Wavelet transform decomposes a signal into a set of basic functions. These basis functions are called wavelets [7].

Wavelets are obtained from a single prototype wavelet $y(t)$ called mother wavelet by dilations and shifting:

$$\varphi_{a,b}(t) = \frac{1}{\sqrt{a}} \varphi\left(\frac{t-b}{a}\right) \quad (3)$$

Where a is the scaling parameter and b is the shifting parameter

1D Wavelet Transform is given by:

$$W_f(a, b) = \int_{-\infty}^{\infty} x(t) \psi_{a,b}(t) dt \quad (4)$$

Its inverse is given by:

$$x(t) = \frac{1}{c} \int_0^{\infty} \int_{-\infty}^{\infty} W_f(a, b) \psi_{a,b}(t) db \frac{da}{a^2} \quad (5)$$

$$\text{Where } C = \int_{-\infty}^{\infty} \frac{|\psi(\omega)|^2}{\omega} d\omega < \infty$$

2.3 QUALITATIVE ANALYSIS

ECG compression literature includes many distortion criteria for performance evaluation. Among the mostly used measures is the PRD [8],[9],[10].

$$PRD = \sqrt{\frac{\sum_{n=1}^N (x(n) - \hat{x}(n))^2}{\sum_{n=1}^N x^2(n)}} \times 100 \quad (6)$$

Where

PRD is Percent Root mean square Difference

$x(n)$ is the original signal

$\hat{x}(n)$ is the reconstructed signal and N is the length of the segment over which is PRD is calculated [11],[12],[13]. In some of the articles, another version of PRD definition is used [11], [12], [13].

$$PRD = \sqrt{\frac{\sum_{n=1}^N (x(n) - \bar{x})^2}{\sum_{n=1}^N (x(n) - \bar{x}(n))^2}} \times 100 \quad (7)$$

Where

\bar{x} is the average value of the original signal.

Equation (6) is dependent on mean value of the original segment; therefore segment with high means will exhibit an artificially low PRD.

Another distortion measure is the signal to noise ratio and is expressed as [14].

$$SNR = 10 \log_{10} \left(\frac{\sum_{n=1}^N (x(n) - \bar{x}(n))^2}{\sum_{n=1}^N (x(n) - \bar{x})^2} \right) \quad (8)$$

Relation between SNR & the PRD (7) is

$$SNR = 40 - 20 \log_{10}(PRD) \text{ dB}$$

Or

$$PRD = 10^{\left(\frac{-SNR}{20}\right)} \times 100 \quad (9)$$

Cross Correlation (CC) measures that determine the amount of similarity between original & reconstructed signal is another objective criterion for testing the quality of reconstructed signal. It is given by

$$CC = \frac{\sum_{n=1}^N (x(n) - \bar{x})(\hat{x}(n) - \bar{\hat{x}})}{\sqrt{\sum_{n=1}^N (x(n) - \bar{x})^2} \sqrt{\sum_{n=1}^N (\hat{x}(n) - \bar{\hat{x}})^2}} \quad (10)$$

3 RESULTS

Different ECG signals from MIT/BIH arrhythmia database were analyzed using MATLAB software. And the various parameters that were used for analysis are PRD, SNR & CC.

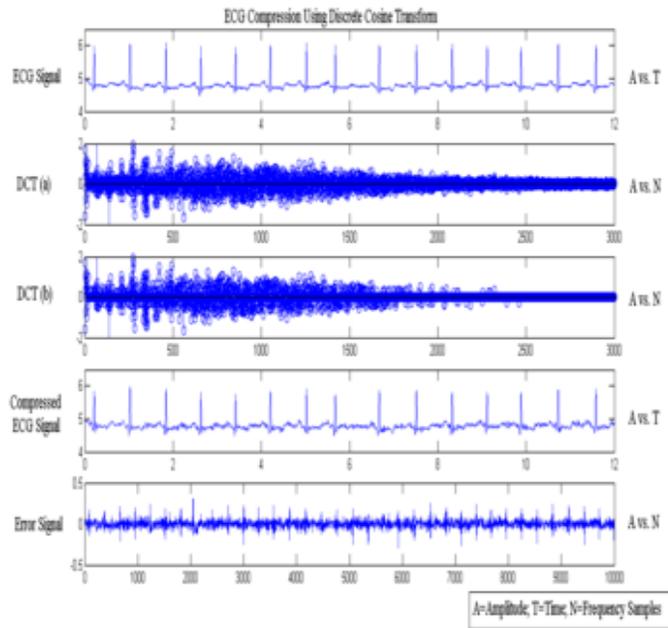


Fig. 1 Simulated result of DCT Compression technique 1(a): Normal ECG Lead II Wave 1(b,c): DCT of ECG Wave by different sampling periods 1(d): IDCT and last one is the error signal.

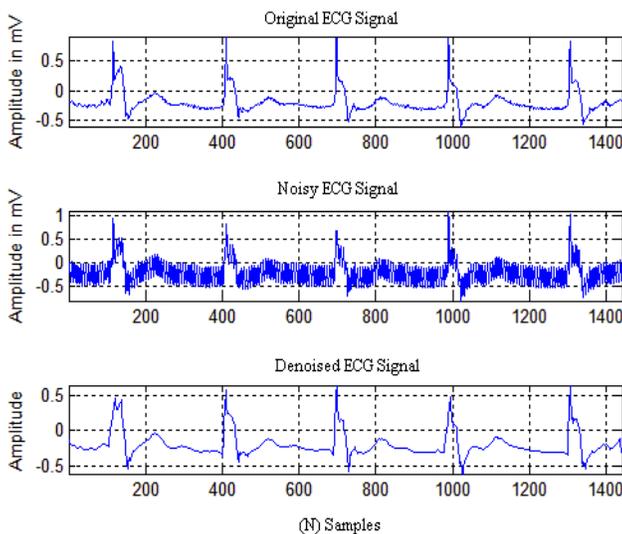


Fig 2 Simulated result of DWT Compression technique 2(a): Original ECG signal 2(b): Contaminated signal while transmission 2(c): Compressed ECG signal

4 CONCLUSION AND OBSERVATION

Electro cardiogram (ECG) is used for the measurement of electrical activity of the heart. In our report we study the important parameters of ECG signal.

ECG signal being used in a wide variety of biomedical applications requires accurate results, less power requirements, faster results and low maintenance cost.

Therefore compression plays a very important role in acquiring these purposes without losing the original information. To evaluate the performance of these algorithms, several objective and subjective criteria are available in ECG literature. Performance evaluation of the considered measure shows that it almost matches the results obtained using subjective evaluation tests. Results indicate that by using the above techniques the use of online data communication schemes can be enabled with high compression without sacrificing the quality of the transmitted signals.

According to our theoretical analysis, and according to what study we have done on the compression techniques, we conclude that DWT compression is highly efficient than DCT compression. DWT gives a signal with the most accurate results after compression and still gives a far better compression ratio than DCT.

Most of the biomedical data compression methods have been developed for ECG signals. However, these methods can be applied to other biomedical signals with some modifications. It is difficult to compare the efficiency of biomedical data compression schemes because coding results of various data compression schemes are obtained under different recording conditions such as sampling frequency, bandwidth, sample precision, and noise level, which may drastically affect the currently used performance measures.

According to our evaluation compression DCT Ratio is 90.430.

5 ACKNOWLEDGMENT

The authors express sense of gratitude to respected Mr. Hemant Kasturiwale, Assistant Professor, Department of Electronics Engineering, Thakur College of Engineering & Technology, for his inspiring guidance and being patient enough to make us understand our mistakes and being tenacious enough all the time to rectify them. His support was admirable all the time till we could get every bit of it. It would have been impossible without his support and guidance all throughout.

6 REFERENCES

- [1] "A Comparison of Single Lead ECG Data Compression Techniques" <http://www.hcs.harvard.edu/~weber/HomePage/Papers/ECGCompression>.
- [2] Laia Bayarri Portolés 2009-09-15 LiTH-IMT-ERASMUS-R-09/34-SE, "Lossless Compression of ECG signals" Performance Analysis in a Wireless Network.
- [3] Victor-Emil Neagoe, Iuliana-Florentina Iatan and Sorin Grunwald, "A Neuro-Fuzzy Approach to Classification of ECG Signals for Ischemic Heart Disease Diagnosis"
- [4] Chia-Hung Lin, "Frequency-domain features for ECG beat discrimination using grey relational analysis-based classifier" *Computers & Mathematics with Applications* archive Vol.55, Issue 4, February 2008.
- [5] J.A. Lay, Ling Guan, "Image retrieval based on energy histograms of the low frequency DCT coefficients," *icassp*, vol.6, pp.3009-3012, Acoustics, Speech, and Signal Processing, 1999. Proceedings vol.6, 1999 IEEE International Conference on, 1999.
- [6] Abstracts of NHLBI-Supported Research, 1972 - Present (links to the NIH CRISP Database)
- [7] Data Compression: The Complete Reference David Solomon – Springer (4th Edition – Dec 2006)
- [8] S.G. Miaou, H.L. Yen, and Ch.L. Lin, "Wavelet based ECG compression using dynamic vector quantization with tree code vectors in single codebook," *IEEE Trans. Biomed. Eng.* vol.49, no.7, pp.671-680, Jul. 2002
- [9] M.L. Hilton, "wavelet and wavelet packet compression of electrocardiograms," *IEEE Trans. Biomed. Eng.* vol.44, no.5, pp.394-402, May 1997.
- [10] Z. Lu, D.Y. Kim and W.a. Pearlman, "Wavelet compression of ECG signals by the set partitioning in hierarchical trees (SPIHT) algorithm," *IEEE Trans. Biomed. Eng.* vol.49, no.7, pp.849-856, Jul. 2002
- [11] A. Alshamali and A.S. Al-Fahoun, "Comments on 'an efficient coding algorithm for the compression of ECG signals using the wavelet transform,'" *IEEE Trans. Biomed. Eng.* vol.50, no.8, pp.1034-1037, Aug. 2003.
- [12] Y. Zigel, A. Cohen, and A. Katz, "ECG signal compression using analysis by synthesis," *IEEE Trans. Biomed. Eng.* vol.47, no.10, pp.1308-1316, Oct. 2000.
- [13] "The weighted diagnostic distortion (WDD) measure for ECG signal compression," *IEEE Trans. Biomed. Eng.* vol.47, no.11, pp.1422-1430, Nov. 2000.
- [14] M. Ishijima, "Fundamentals of the decision of optimum factors in the ECG data compression," *IEEE Trans. Inf. Sys.* vol. E76-D, no.12, Dec. 1993.

-
- Nikhil Dhawan is currently pursuing degree program in electronic engineering in Thakur College of Engineering and Technology from University of Mumbai, India, PH- +919029495025.
E-mail: nikhil.h.dhawan@gmail.com
 - Saurabh Gupta is currently pursuing degree program in electronic engineering in Thakur College of Engineering and Technology from University of Mumbai, India, PH- +919820449145.
E-mail: 100rabhgupta11@gmail.com
 - Ravi Sadrani is currently pursuing degree program in electronic engineering in Thakur College of Engineering and Technology from University of Mumbai, India, PH- +919870073052.
E-mail: ravi.sd101@gmail.com
 - Swati Sanghvi is currently pursuing degree program in electronic engineering in Thakur College of Engineering and Technology from University of Mumbai, India, PH- +919773240450.
E-mail: swati13sanghvi@gmail.com

A new method for quantitative evaluation of diamond wheel sharpness in EDDG of advanced cutting tool material

Siddharth Choudhary, Pratik Jain and Dr. S.K. Jha

Abstract: In this paper authors propose a probability statistic based method for evaluating metal bonded diamond grinding wheel sharpness quantitatively. Grinding performances are decided by wheel sharpness and the conditions under which it is sharpened. It is also characterized grinding forces, cutting zone temperatures, power consumption and surface finish of components. Many times, in shop floors, the diamond wheels are used in a very inefficient manner due to lack of proper information about the wheel sharpness.

The correlation of wheel topography and its performance is studied through the employment of three dimensional surface characterization parameters. Wheel sharpness is analyzed by using the probability statistics in order to ascertain quantitative measurements of the sharpness of the wheel face. Wheel sharpness are evaluated by considering the influences of impulse energy, cutting speed, cross feed (depth of cut), and grit size on the same. Electrical discharge diamond grinding (EDDG) plays an important role in realizing highly efficient and accurate grinding of extremely hard electrically conductive tools materials. Electrical discharge diamond grinding, i.e., grinding with metal bonded diamond wheel under the , has been successfully applied in many instances since the early 19 Century for making cutting tools. EDDG is the most suitable process for manufacturing good quality cutting tools made of very hard electrically conductive tool material. Ductile mode grinding of brittle materials has been of great interest due to its increasing industrial applications and academic demands for fundamental understanding of the ductile mode grinding mechanism.

Keywords: Electrical discharge diamond grinding, Metal bonded diamond grinding wheel

◆

INTRODUCTION

Recently, developments in the cutting tool industry have caused rapid increase in the use of advanced tool materials to distinctive progress. However, this has meant that grinding of these materials to form the cutting tools itself has also, progressively, become more difficult [1, 2]. As a solution to this difficulty application of hybrid machining processes has been the main trend observed in the cutting tool industry in need of advanced tool materials. The concept of electrical discharge diamond grinding (EDDG) which originated in the former USSR, is one of such processes used extensively in the manufacture and maintenance of the cutting tools made of in particular when they are tipped with electrically conductive very hard materials. A metal bonded diamond grinding wheel (MBDGW) is used in this process. MBDGW is used to remove material by mechanical action, and also acts as an electrode for electro discharge action, causing thermal softening of the material and hence machining by diamond grains become easier. EDDG allows the use of water or water based cutting fluid as a dielectric. Wheel speed and electrodischarge pulse parameters are adjusted to control the mechanical grinding and electrodischarge erosion during EDDG. As a result, the grinding wheel is constantly maintained in optimum state of sharpness or adjusted to the processing operation currently in hand. In EDDG, by

adjusting wheel speed and electrodischarge pulse parameters, sharpness of the topography of the wheel can be adjusted to the machining task in hand.

However, a technical basis to utilize the diamond wheel to its best potential and in an economical manner require evaluation of diamond wheel's sharpness at the microscopic level ,i.e., from the considerations of grain work interaction which, in turn, will allow predicting wheel performance more accurately [7]. However, a methodology is required to allow the wheel sharpness to be used at its best potential. Clearly, the more realistic the wheel sharpness evaluation method, the more accurately EDDG performance may be predicted.

A number of methods for evaluating grinding wheel sharpness have been developed so far [3]

However, most of the suggested methods are laboratory based requiring standard equipment and cutting conditions. As a result, generalization or estimations of the grinding wheel sharpness are difficult to make with any degree of accuracy and to apply them to industry due to the limitation of equipment technology, staff and methodology. Hence, the term wheel sharpness is so far understood more in qualitative than in quantitative terms. This is more apt in industry where wheel sharpness has been specified by the operators based on their observations and experience. This

empirical approach typically results in a highly variable situation as well as the tendency for too much dressing of the wheel in order to be on the "safe" side. This causes material waste and unnecessarily reduces the wheel's life. Therefore, it is necessary to develop a methodology in order to ascertain quantitative measurement of wheel sharpness.

Actual wheel sharpness can be considered the summation of the sharpness of individual grains/cutting edges [6]. Since the size of cutting edges is inherently random in nature, the wheel sharpness cannot be evaluated in a deterministic manner and one has to resort to a probabilistic approach. Therefore, the aim of the present work was to estimate MBDGW sharpness quantitatively using probability statistic. In addition, this work also evaluates influence of impulse energy, cutting speed, cross feed (depth of cut) and grit sizes on the wheel sharpness. Machining of advanced tool materials to form the cutting tools generally necessitates grinding with diamond wheels. Three flaring cup type wheels with diamond grit in 160/125, 180/190 and 50/40 Russian mesh size respectively were used. Diamond concentration was 100 in all cases. The workpiece material chosen was Tolam-10 (Russian made) type cutting tool material. The result of this work can contribute to the analysis of EDDG performance, optimization of diamond wheel structures and optimization of diamond spark grinding process.

In the grinding of advanced materials such as optical glass, WC, ceramics and silicon has substantially grown with the widespread use of precision components made of such materials in various applications [4, 5].

Considering efficiency and precision, the abrasive grain protrusion height is a crucial parameter of grinding wheel's surface topography to get a high amount of total cutting. While, it is difficult get the protrusion height of an entire surface since there are three elements coexisting with their own geometrical shapes - depth, radius distance and azimuthal angle of flare type grinding wheel.

Concept of the method for evaluating wheel sharpness

In this version of wheel sharpness evaluation method, variations in diamond grain's protrusion heights from the wheel surface in the longitudinal direction-the grinding direction is estimated. When it is noted that an individual grain could have many tiny cutting points, then the variations in the heights referred to the level difference of such cutting points of the grain which produces atleast partial cutting above the similar cutting point of the

previous active grain [8]. In fig.1, it may be seen that in the section $Y=Y_j$, Δij is the level difference of one cutting edge of i-grain, which depends, among other parameters, on the applied normal force. Also, it may be seen that among all the cutting edges, only those denoted by α , β , gamma are assumed to be involved in material removal. It may be seen that Δij is one of the fractions of chip thickness or one of the numerous coordinates of the cross sectional area of the chip generated by a cutting edge of i-grain in the given section. Thus, the sum of all the Δij , obtained in various sections i.e. when $j=1\dots P$, is the cross sectional area of the chip, S_i , generated by all the cutting edges of i-grain. It seems, that a wheel with a well sharpened face would have the close values of S_i for different grains, because it contributes to uniform load distribution on the cutting edges, uniform wear characteristics of these edges etc. Therefore, as a measurement of wheel sharpness, the degree of S_i values scattering for different grains, can be used; the less the better. Based on this idea, a wheel sharpness index, originally used for theoretical analysis of the variation in diamond grit heights on the wheel surface, is applied to measure the diamond grinding wheel sharpness in DSG. The following major assumptions are made to generate diamond wheel cutting surface in order to measure the wheel sharpness.

Following this, using a contact stylus type system, the grit protrusion heights are measured and quantitatively described by an appropriate probability density function.

Description and Formulation of the Wheel Sharpness Index

In this version of wheel sharpness evaluation method, variation in heights/separation distance between the surfaces of protruding diamond grains in the longitudinal direction-the grinding direction, is estimated. When it's noted that a grain could have more than one cutting edge, then the variation in the heights refers to the level difference of such cutting edge of the grain which produces atleast partial cutting above the similar cutting edge of the previous active grain. In fig.1, it may be seen that in the section $Y=Y_j$, Δij is the level difference of one cutting edge of i-grain, which depends, among other parameters, on the applied normal force. Also, it may be seen that among all the cutting edges, only those denoted by α , β , gamma are assumed to be involved in material removal. It may be seen that Δij is one of the fractions of chip thickness or one of the numerous coordinates of the cross sectional area of the chip generated by a cutting edge of i-grain in the given section. Thus, the sum of all the Δij , obtained in various sections i.e. when $j=1\dots P$, is the cross sectional area of the chip, S_i , generated by all the cutting edges of i-grain. It seems, that a wheel with a well sharpened face would have the close

values of S_i for different grains, because it contributes to uniform load distribution on the cutting edges, uniform wear characteristics of these edges etc.

Hence, if we denote selected average values of S_i for all grains on the chosen surface as S , and selected deviation of S_i values as σ , then the wheel maybe characterized quantitatively as function $W_s = f(S, \sigma)$. It is quite obvious that the above function should be increasing with respect to the first argument S and decreasing with respect to the second argument σ respectively then, ratio S/σ . Therefore, as a measurement of wheel sharpness, the degree of S_i values scattering for different grains can be used; the lesser the better. Based on this idea, a wheel sharp index, originally used for theoretical analysis of the variation in diamond grit heights on the wheel surface, is applied to measure the diamond grinding wheel sharpness in DSG. The following major assumptions are made to generate diamond wheel cutting surface in order to measure the wheel sharpness:

1. The diamond grain is ordinary sharp tip cone with semi cone angle, θ , as shown in fig.2. This geometry is consistent with observation showing abrasive particles to have negative rake angles. The difference in heights Δij is the depth of grain indentation.
2. Protruding heights on the wheel surface follow a gamma distribution. This distribution is consistent with the observations showing that during grinding of difficult to grind material, due to severe thermomechanical cycle, microspalling of the grain, takes place. Therefore, the topography of the wheel surface keeps on changing during "run-in" of the wheel. In view of this, the gamma probability density function is found to be more appropriate in order to describe the protruding heights of the diamond grains on the wheel surface.
3. Since only a portion of whole grain participates in the fine grinding process, an individual grain and its microcutting edges are symmetrical in shape.
4. The removal of the material from the workpiece is based on a rigid perfectly plastic material behavior i.e. the cutting edges will remove all materials that they encounter on their path. Although the elastic deflection of the workpiece will reduce the actual cutting depth (Δij), the possible pile up of material and the deflections are neglected.
5. Diamond grains along the circumferential and axial direction of the grinding wheel are distributed uniformly. This is based on the fact that during the manufacturing of diamond wheels, diamond grains, metal bond and the

proper proportion of the filler are mixed thoroughly and the mixture is fed in to revolving drums, where all constituents mix together to form a paste. This process provides desired consistency to the mixture coats the diamond grains with adequate bond and produces uniform spacing between them. The paste is then placed in moulds to get the shape of the wheel.

6. There is no interaction between the bond and the workpiece. This is based on the fact that the DSG is able to adjust and maintain a constant grit protrusion and a stable grinding performance.

Based on these assumptions the relationship between wheel sharpness w_s the selected average value of all chip cross sectional areas generated by the grains on the wheel surface S and selected standard deviation of these areas σ , is approximated as

$$W_s = S / \sigma$$

Techniques for calculating S and σ values are described in the next section.

3. Experimental Validation

3.1 Materials and Experimental

The adaption of ECCD technology to the universal tool and cutter grinder, model 3B642 (Russian made) was essentially achieved by fixing brushes in the grinding spindle head. By these brushes the power was supplied to the rotating grinding wheel (anode). The current source was connected to the brushes and the workpiece (cathode) respectively was achieved by brushes, which were fixed in the grinding spindle head. For this purpose, grinder spindle was modified for the supply of current into the spindle from which it then flowed to the rotating grinding wheel. As a power source, specially fabricated small sized generator was used whose characteristics could be controlled to provide optimum electrical parameters for EDDG of Tomal-10. As a dielectric fluid, tap water with small amount of soda was used. The materials of the workpiece were type-1 and type-2 carbides. The composition and the properties of these carbides are given in Table 1 and 2 respectively. For the experimentation bronze brass alloy bonded diamond wheels (flaring cup type) with 100% diamond concentration and the grit size 100/80 (USSR mesh) were used. These wheels and the grinding conditions shown in table -3 were selected in accordance with the experimental study on finding optimum grinding condition for EDDG of type-1 and type-2 carbides [11]. After grinding the workpieces, worn wheel faces were photographed with

help of Scanning Electron Microscope (SEM), Model jeol-JSM-50.

The experimental work for the quantitative measurement wheel sharpness can be divided into 2 parts. In the first part, the grit protrusion on the wheel surface is measured using a suitable wheel topography measurement system. The statistical parameters derived from such measurements are given in the form table and also a cumulative distribution function which is then approximated by a standard distribution function. Then the parameters for wheel cutting surface generation are determined. Using these parameters, generation of diamond grains on certain area of the wheel cutting surface is performed. Finally, variation in heights of diamond grains, S and σ values and then wheel sharpness index are calculated. In the second part, effect of some DSG parameters on the wheel sharpness index is analyzed.

3.2 Measurement of Protrusive Height of Diamond Grits Above the Bond

The protrusive heights of the diamond grits on the surface of three metal bond flaring cup wheels containing diamond grit in 160/125,180/90 and 50/40 Russian mesh size respectively and 100% diamond concentration in all cases were measured using specially designed stylus-based measurement system, also called as stylus profilometer. The diamond stylus was traversed on the surface to be measured over a predetermined distance of 15mm, 15mm, and 30mm for and diamond grain on the profile chart, a hard metal stylus of suitable size was mounted near Tomal-10 made indexable inserts under the conditions shown in table. Composition and mechanical properties at ambient temperature of workpiece material is given in table.

Fig shows the typical cutting surface of the wheel measured using stylus profilometer. The cumulative distribution functions of diamond grains protruding heights derived from such measurements on wheel surfaces is shown in fig. Its seen clearly from fig and that the diamond grain protruding heights of these three wheels can be approximated by a bounded gamma distribution function. Similar observations have been made by Katsev. The statistical parameters derived from this experiment are listed in the table

3.3 Parameters for wheel cutting generation

From the point of view of wheel cutting surface generation, the following parameters were used:

3.3.1 Mathematical equation for the diamond grain shape

The basic conical shape of the diamond grain can mathematically be formulated as follows:

$$Z = Z_0 - \theta [(X - X_0)^2 + (Y - Y_0)^2]^{1/2} \text{ -----(1)}$$

where Z of the height of the grain, θ is the tangent of semi-cone angle of the cone shaped grit which is calculated by (Refer fig...)

$$\theta = \tan^{-1} (R/h)$$

where, R is the radius of the cone shaped grain and h is the grain height. X, Y, Z are the coordinates of cone summit.

3.3.2 Diamond grains protruding height distribution on the wheel surface

The height from the bond surface to the very top of the diamond grit, δ , is not the same for all the diamond grains on the wheel surface but are statistically distributed between a smallest value, and a larger value. Generally, this distribution can be well described by an appropriate bounded probability function. In this work, gamma probability function is used. This function is expressed mathematically as

$$f(\delta) = (\beta \alpha + 1) \cdot e^{-\beta \delta} \cdot \delta^{\alpha} / (\alpha + 1)!$$

On the basis of calculated values of mathematical expectation and variance, α and β parameters for different grain sizes were calculated as

$$\alpha = 6.058 - 0.083 Z + 0.00076 Z^2$$

$$\beta = 0.555 - 0.0056 Z + 0.00003 Z^2$$

where Z is the grain size which is the mean diameter of the grain.

3.3.3 Diamond grains mean diameter

The mean diameter d_{mean} , of the grain was determined as

$$d_{mean} = (d_{max} + d_{min}) / 2$$

where d_{max} is diameter of sieve through which the grain passes and d_{min} is that on which grain remains. Table 6 gives values of mean grain diameters for different grain sizes.

3.3.4 Diamond grains distribution along X and Y directions

It has been assumed that the spacing's between two consecutive grains along X and Y directions follow uniform distributions. These distributions are shown in Fig...

3.3.5 Number of diamond grains on the wheel surface

The number of abrasive grains on the grinding wheel surface was determined by counting the cutting points on the profile trace obtained from the profilometer. The number of grains on certain cutting surface area of the wheels with different grain sizes is given in table 6.

Function corresponding to the probability density function f (delta). Thus the location of the grains on the wheels surface is decided. Figure shows the simulated wheel cutting surface.

3.4 Determination of variation in heights of diamond grains

Once the aforementioned parameters are known, random number generator is used to simulate the positions of diamond grits referring to the principles of the Monte Carlo method. After the diamond grains on an unit cell of certain size of the wheel surface are generated, as schematically as shown in fig The coordinate system XYZ is fixed to that unit cell at point O. Let the unit cell be divided into n equally spaced divisions along the Y direction. The line m at the distance Ymax/2 denotes one of such divisions that intersect the cutting points of N number of diamond grains. Then, the protrusive height, Δ of all these cutting points are determined using the formula

$$\text{deltai} = Zi - a [(m - Yi)]^{1/2}$$

As a result, a set of delta1,2,3,4.....n values are obtained. On the basis of these values, variation in heights of the cutting points for each grain in the given division, Δ ij are determined. Similarly, Δ ij for other divisions are calculated. As a result, a set of Δ ij values for various cutting edges of an individual grain are obtained. By integrating the values, the cross sectional area of the chip, Si, generated by all the cutting edges of an individual grain is obtained. In the present work trapezium method is used for integrating the Δij values. Similarly, Si for other grains are calculated which when integrated, the cross sectional area generated by all the grains, S is obtained.

3.5 Calculation of S and σ values

When the sum of cross sectional areas generated by all the grains on the simulated wheel surface , S, is known the mean cross sectional area can be determined as

$$S' = S/n = (S1 + S2 + S3 +.....+ Sn)/n$$

and the standard deviation of the cross sectional areas from the mean value , σ, is determined as

$$\sigma = \{[(S1 - S')^2 + (S2 - S')^2 +.....+ (Sn - S')^2] / n\}^{1/2}$$

where n is the number of diamond grains on the simulated area of the wheel surface. The calculation and also the assessment is carried out by computer and the results obtained can be displayed as graphs and can also be output as text file. The flow chart for calculating the wheel sharpness is shown in fig .Table below gives values of S, σ and Ws in tabular form for convenience.

3.6 Effect of DSG process parameters on wheel sharpness index

To assess the effect of DSG process parameters on wheel sharpness index, grinding tests were carried out on Tomal - 10. The workpiece was ground with wheel that had been sharpened by DSG process. The test conditions are entered in the corresponding diagrams.

3.6.1 Effect of single discharge energy on the wheel sharpness index

As maybe seen in fig 2, with the decrease in single spark energy, wheel sharpness index initially increases and after that decreases. The maximum Ws is when the pulse frequency, f = 44 KHz. At f = 22KHz, intensified sharpening takes place. However, greater spread in the values of variation in heights of the cutting points, have an influence on σ. Decrease in Ws at f = 66KHz can be explained by the fact that at small discharge energy (6 X 10-4 Joule), active self sharpening of the wheel surface does not take place. Consequently, difference in the height of protrusion of all the tips of the grains becomes much reduced. As a result, the mean cross sectional area generated by the grains is less. Therefore the sharpness index has lower value. The above results show that when grinding Tomal-10, the pulse frequency, f= 88KHz be employed in order to maintain the wheel sharpness at high level.

3.6.2 Effect of the wheel speed on sharpness index

The effect of the wheel speed on the sharpness index during the DSG of Tomal-10 is shown in fig 3. The wheel sharpness index diminishes as wheel speed increases. There are various reasons for the shape of this curve. The higher speeds lead to reduced cutting cross sections with the result that the load on the grains, but also the removal per grain

are reduced. Neither the increasing wheel speed nor the higher engagement frequency of the diamond grains on the diamond layer can compensate for the diminishing removal and the wheel sharpness index as well. In addition, the sliding effect is intensified by the higher speeds. These mechanisms lead to reduction of removal at higher speed.

3.6.3 Effect of the cross feed on the sharpness index

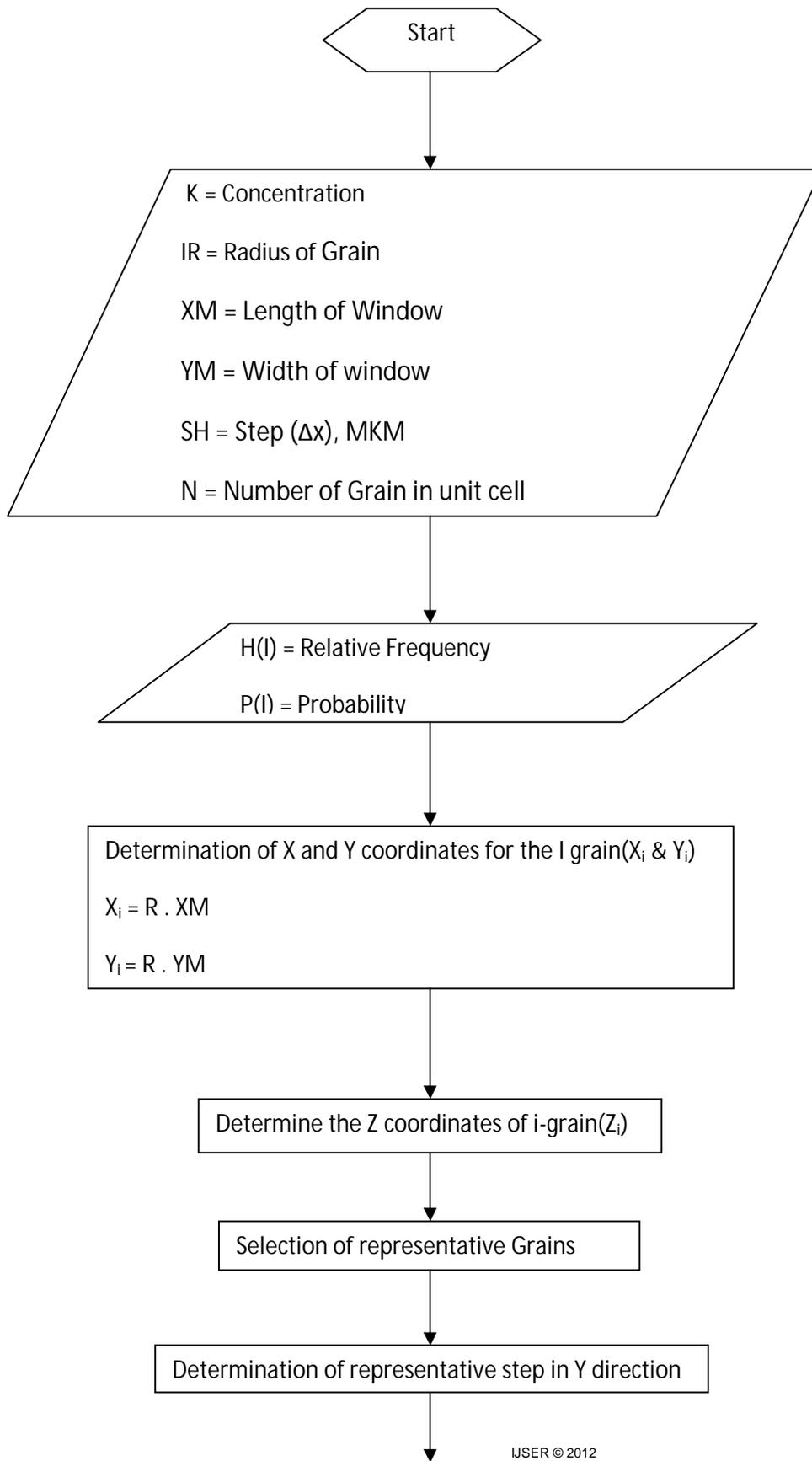
The sharpness index of the diamond wheel with grit size 100/80 is plotted by function of the cross feed in fig 4 below. The WS increases with increase in cross feed. This can be explained by the fact that, when the workpiece is hard, at a small cross feed only some of the diamond grains penetrate the material, most of them only slide and thus remove less material. The grain tips are correspondingly blunted on the workpiece and thus causes glazing of the wheel. Consequently, cross sectional area generated by the grains becomes small and therefore the wheel sharpness index decreases at small cross feed. Penetrate with increase in cross feed, the diamond grains penetrate farther into the material and thus remove more for each grain engagement. As a result, cross sectional area generated by the grains increases and therefore the wheel sharpness index also increases.

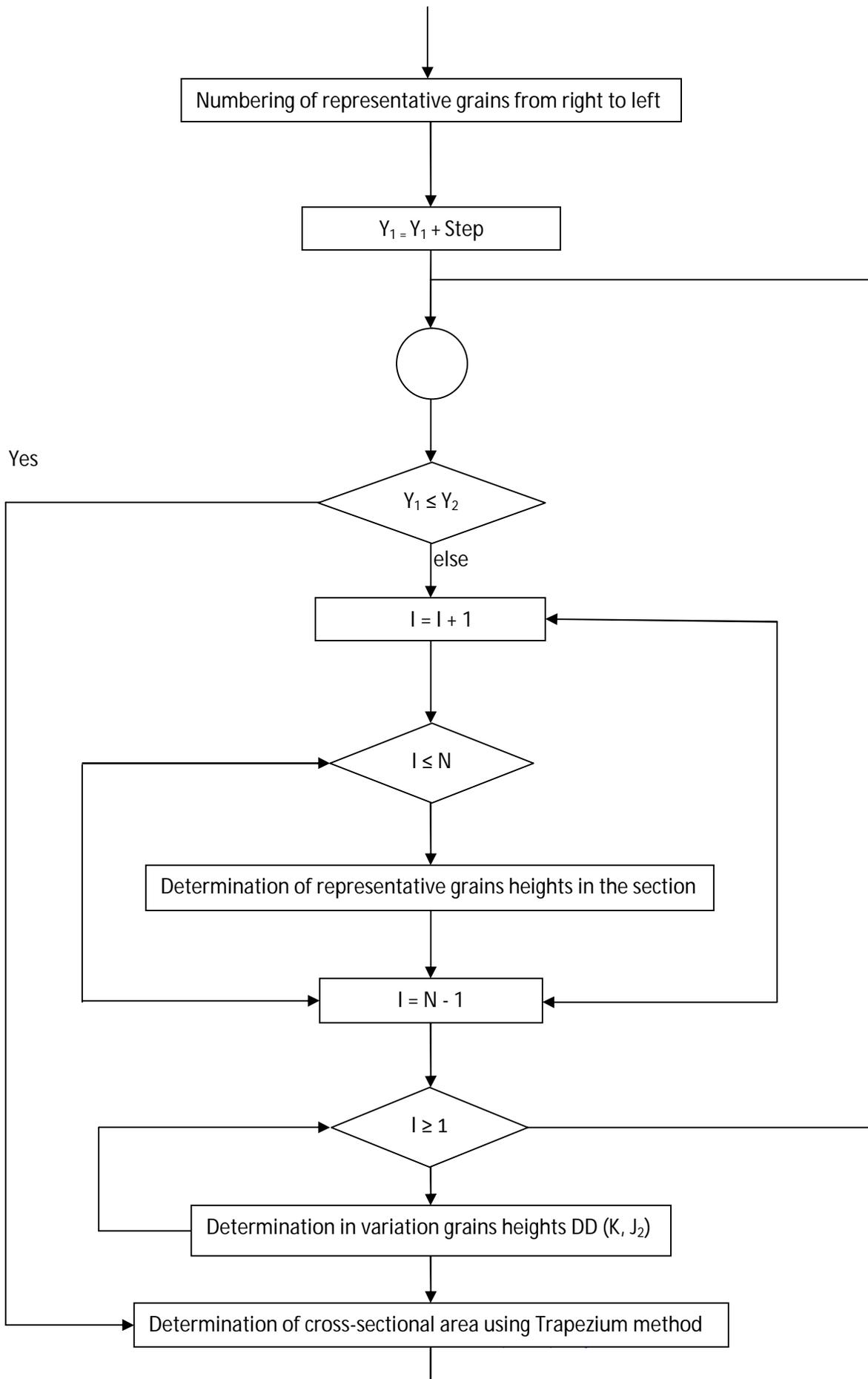
3.6.4 Effect of the grit size on sharpness index

Diamond wheels with fine grit size are normally used for finishing Tomal-10. A comparison of the wheel sharpness indices with 50/40, 180/90 and 160/125 Russian mesh wheels reveals no significant difference in the wheel sharpness

indices for 180/90 and 160/125 wheels, whereas the sharpness index is lowest for the finer grained (50/40) wheel (fig 5). Despite the increase in the number of active grains with decreasing grit size, the removal rate does not, however, achieve the values achieved with the coarser grained wheel which is attributed to the larger sliding area that counteracts the increase in material removal rate, i.e., with a fine wheel size, the workpiece rests on several tips and slides. With coarser grain sizes, the grains penetrate further into the material and thus cause the slightly higher tangential forces, with the result that the material removal also increases.

The comparative tests with the above wheels show that Tomal-10 should be ground with coarser grained diamond wheels.





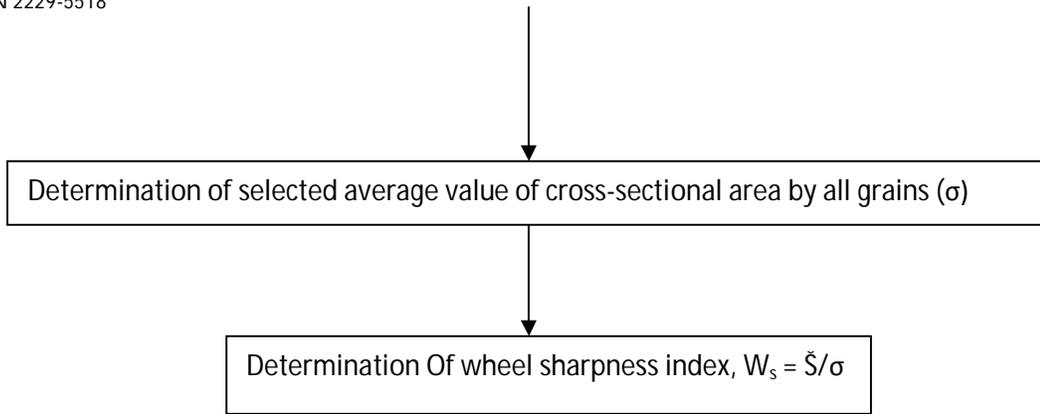
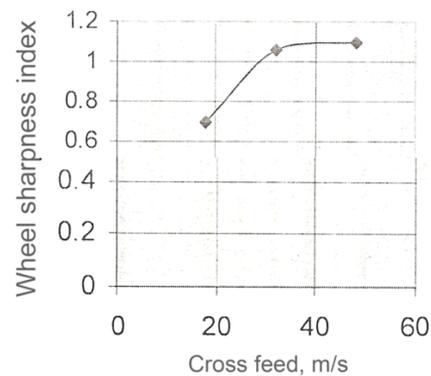
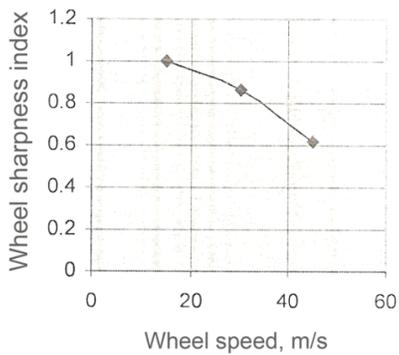
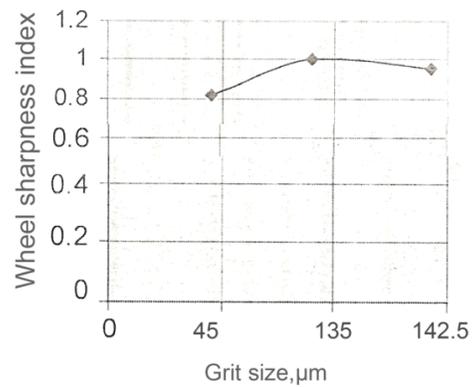
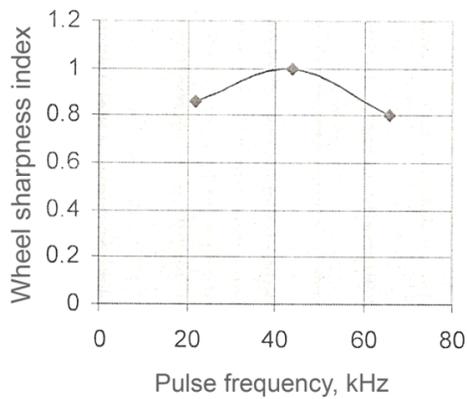


Fig.1 Flow chart for calculating the Grinding Wheel Sharpness

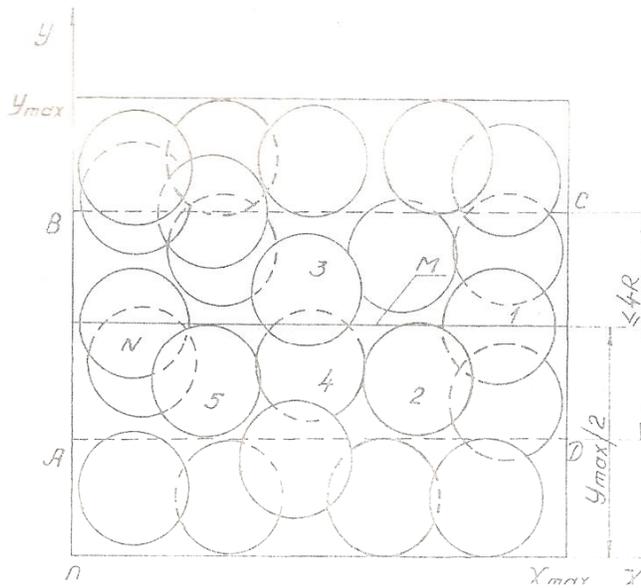
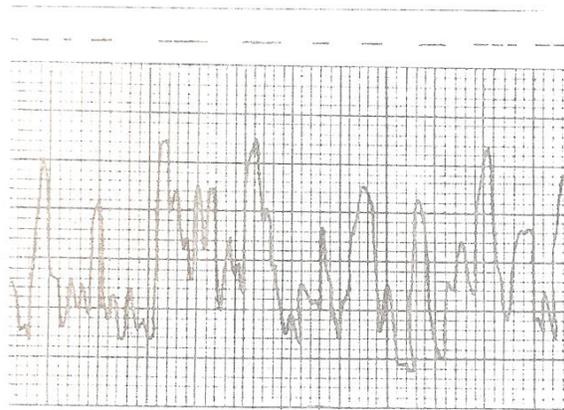


(starting from left) Fig 2 : Relationship between the pulse frequency and wheel sharpness index

Fig 3: Relationship between the wheel speed and wheel sharpness index

Fig4: Relationship between the grit size and wheel sharpness index

Fig 5: Relationship between the cross feed and wheel sharpness index



Wheel grit size (Russian mesh)	50/40				100/80				160/125			
	I	II	III	Mean	I	II	III	Mean	I	II	III	Mean
Modeling Variant												
Area of the unit cell in consideration (mm ²)	0.86	0.83	0.83	0.84	1.76	1.79	1.79	1.78	5.49	5.55	5.49	5.51
Number of grains selected on the unit cell	22	29	29	27	23	26	20	24	32	24	23	26
Number of cross sectional areas generated by the grains on the unit cell	11	8	8	9	14	13	10	12	8	11	9	9
Integrated cross sectional area generated by individual cutting grains (mm ² x 10 ⁻⁶)	49.5	45.5	45.5	46.83	86.2	87.5	89.3	87.2	100.1	115.4	132.5	116
Mean of the cross sectional areas generated by all grains (mm ² x 10 ⁻⁶)	4.5	6.18	6.18	5.62	7.1	7.4	6.8	7.1	14	16.2	17.5	15.9
Standard deviation of cross sectional areas from the mean value (mm ² x 10 ⁻⁶)	3.85	8.89	8.89	7.21	6.72	6.56	6.62	6.63	8.89	9.31	10.09	9.7
Wheel sharpness index, W _s (e ⁻¹⁵)	0.59	0.69	0.69	0.65	0.82	0.84	0.87	0.83	1.57	1.62	1.69	1.62

Conclusion:

The above discussion demonstrates that the following conclusions:
Theoretical measurement of particle and surface sharpness has been developed.
The protruding height of each grain is measured and wheel sharpness has been evaluated and demonstrated in a tabular form.
The effects of single discharge energy, grit size, cross feed, DSG process parameters, wheel speed were studied.

References

[1] S. Malkin, Grinding Technology—Theory and Applications of Machining with

Abrasives, Ellis Horwood Ltd., Chichester, UK, 1989.
 [2] M.C. Shaw, Principles of Abrasive Processing, Oxford University Press, Oxford, 1996.
 [3] S. Malkin, Grinding Technology, Ellis Horwood, Chichester, UK, 1998.
 [4] S. Agarwal, P Rao Venkateswara, A new surface roughness prediction model for ceramic grinding. Public roc inst Mech Eng, B J Eng, Manuf 219 (11): 811 – 821, 2005.
 [5] H.T. Young, H.T. Liao, H.T. Huang Surface integrity of silicon wafers in ultra precision machining. Int J Adv Manuf Technol 29 (3 - 4): 372 - 378, 2006.

- [6] J. Verkerk, Final report concerning CIRP cooperative work on the characterization of grinding wheel topography. Ann. CIRP 26 (2), 385–395, 1977.
- [7] K.J. Stout , P.J. Sullivan, W.P. Dong , E. Mainsah, N. Luo, T. Mathia, H. Zahouani, The Development of Methods for the Characterisation of Roughness in Three Dimensions. EC, Luxembourg, 1993.
- [8] American Society of Mechanical Engineers, ASME B64.1-1995 Surface texture: surface roughness, waviness, lay, 1995.
- [9] J. Goddard, H. Wilman, A theory of friction and wear during the abrasion of metals, Wear 5 (1962) 114–135.
- [10] D.V. De Pellegrin, G.W. Stachowiak, A new technique for measuring particle angularity using cone fit analysis, Wear 246 (2001) 109–119.
- [11] H. Sin, N. Saka, N.P. Suh, Abrasive wear mechanism and the grit size effect, Wear 55 (1979) 163–170.
- [12] D.V. Pellegrin, G.W. Stachowiak, Assessing the role of particle shape and scale in abrasion using ‘sharpness analysis’. Part II. Technique evaluation, Wear 253 (2002) 1026–1034.
- [13] W.H.M. Robins, The significance and application of shape factors in particle size analysis, in: H.R. Lang (Ed.), The Physics of Particle Size Analysis, The Institute of Physics, London, 1954.
- [14] L. Blunt, S. Ebdon, The application of three-dimensional surface measurement techniques to characterising grinding wheel topography, Int. J. Mach. Tools Manuf. 36 (11) (1996) 1207–1226

The Determinants of Migration and Remittances in Albania

Esmeralda Shehaj

Abstract— This paper uses the recent Albanian Living Standards Measurement Survey 2008 to analyze the determinants of migration and remittances in Albania. It addresses one of the main limitations of the literature on migration and remittances, namely their separate study, and analyzes the determinants of these two phenomena jointly. The analysis is focused on the household, and the migration decision process is made by the household as a whole, seeking to maximize expected future utility, which is achieved through remittances sent by the migrant. In terms of determinants of migration and remittances the study shows that migration and the receipt of remittances are selective processes strongly affected by household characteristics. We do not find evidence for the existence of a 'migration hump' with respect to per capita expenditures and/or household wealth, and the brain drain does not seem to be an issue. The probability to migrate and remit of at least one household member is found to be influenced mainly by household characteristics number of adults and number of children, gender ratio and location. Unlike many other studies we control for relative deprivation, and find evidence in favour of the relative deprivation theory of migration.

Index Terms— Albania, Migration, Remittances, Relative deprivation, Selectivity, Social capital.

1 INTRODUCTION

USING the data from the Albanian LSMS 2008 this study develops and estimates a model to investigate the determinants of international migration and remittances in Albania. There is a vast of research trying to explain migration flows, individual migration and remittances using different econometric approaches and explanatory variables. Our empirical analysis has the household at its focus, and for this reason we review the microeconomic literature, concentrating on the household level empirical determinants of migration and remittances. The approach followed in this study enables testing for more than one theory at a time and helps to identify some sets of common or widely used explanatory variables at the household level in order to delineate the factors that influence the decision to migrate and the receipt of remittances. During the last decades the empirical literature on the motivations to migrate and remit has grown, but the results have often been conflicting. While these inconsistent results may be attributed to differences in the context and characteristics of the country under consideration, the empirical approach, or data availability, one common shortcoming is that they are usually based on testing particular theoretical models of migration and/or remittance receipt. Recent research has shown that none of the theories of migration alone can explain all the dynamics of migration and receipt of remittances, the motives may overlap and it is very difficult to disentangle and they are not exclusive. In the last decade a few studies on the determinants of remittances have incorporated factors to ameliorate the limitations of these theories.

The conventional approach of the empirical modelling strategies treats migration and remittance behaviour as independent decisions. Other approaches consider the decision to migrate

and send remittances back home as interrelated in different ways, which is argued to be more appropriate. Focusing exclusively at the determinants of remittances and omitting the importance of factors that influenced the migration decision may not only leave out these crucial factors, but also bias the results. Thus, linking both decisions would yield to more accurate determinants of remittances. Considering migration and remittance behaviour as interrelated decisions is also arguably more appropriate empirically. First, it allows controlling for the possibility of endogeneity of the two decisions, considering the decision to remit as an important determinant of migration itself. Second, it also enables modelling migration as a selection mechanism for remittances, correcting for the selection-bias of the estimates. This second possibility gains even more importance when differentiating between the desire and the capacity to remit. Following these arguments, and based on the similarities of the sets of variables from different conventional approach studies when explaining migration or remittances, we consider a model that explains the joint phenomena of migration and remittances at the household level.

The paper is organized as follows. Section 2 presents a literature review of the Albanian research in the field. Section 3 includes the empirical approach that will be followed, the data set, the empirical variables that will be used and their measurement. The focus of Section 4 is the estimation of the empirical model and the interpretations of the results, and the last section concludes.

2 THE DETERMINANTS OF MIGRATION AND REMITTANCES IN ALBANIA

Carletto et al. (2004) studied the determinants of temporary and permanent migration from Albania to different countries. The household characteristics include family size, age of the head of the household, demographic composition, average adult education, agricultural assets (land and livestock), la-

• The author is currently pursuing doctoral degree program in economics at the Staffordshire University, United Kingdom.
E-mail: Esmeralda.alb@gmail.com

bour activities and wealth proxies (previous ownership of a vehicle and the number of rooms per capita). They argue that limiting the number of assets to be included in the model would minimize the endogeneity problem. Actually, this may create another endogeneity problem, that of the omitted variable, which will also yield biased and inconsistent estimates. Their estimates suggest that most permanent migrants are young males, who come from larger households, with an older head of household and fewer smaller children. Education is not an important determinant of migration which may be attributed to the fact that most Albanians have finished secondary school.

The type of labour activities seems to be an important determinant of the destination country and migration duration; ownership of cattle is negatively associated with both temporary and permanent migration; the existence of migration networks and previous experience with migration are key determinants in the decision to migrate internationally, while community level networks are important only for temporary migration. Relative wealth is also a factor in the decision to migrate with the deprivation of a household relative to other households at the village level positively associated with the decision to migrate.

Finally, regional factors play a role in the migration decision. Households living in Tirana are less likely to migrate internationally. This is particularly true for permanent migration, in which case households living in all other regions have a greater probability of migrating than those in Tirana. Compared to Tirana, households in the rural Centre, Coast and Mountain are more likely to migrate temporarily, and households in the urban Coast and Mountain regions to migrate permanently. One possible reason could be that they are already internal migrants, but taking into account that the internal migration is strongest towards poorest peri-urban areas of Tirana, another explanation may be that many of them cannot afford to migrate internationally.

Konica and Filer (2009) use a migration survey of 1000 households carried out in 1996 to study the determinants of migration and amounts of remittances. The explanatory variables in the Probit equation of migration are the individual characteristics of the migrants, geographic indicators, and two household level variables: income and size of the households. The results indicate that large, rural, and low-income households are more likely to send someone abroad. At the individual level, young, male, single, high school graduates and the unemployed are more likely to migrate.

With regard to amounts of remittances sent to households in Albania their results indicate that remittances are positively related to the employment status of the emigrant, the presence of a spouse in Albania, the emigrant's legal status, as well as whether the emigrant had arranged a job in the foreign country prior to departure. The existence of other emigrants from the household negatively influences the amount of cash remitted, while gender and length of stay do not appear to affect remittances.

Lianos and Cavoundis (2004) also surveyed legal Albanian migrants in Greece, and found that women, married migrants, those with family left behind, and those with more relatively

deprived households are more likely to remit. The amounts of remittances are positively influenced by income, less stable employment and number of children in Albania. Germenji, Beka and Sarris (2001) carried out a study to test whether remittances sent to rural households in Albania were sent for altruism or exchange motives. Based in a survey of 200 rural households conducted in 2000, they conclude that most of the variables they control for are insignificant. The only significant variables are the pre-transfer income of household, the presence of more than one member abroad and when the decision to migrate has been seen as a need to help the families. The lack of other significant variables is possible because of the relatively low sample size, as well as limited other explanatory variables. It is not clear whether data limitations are due to questionnaire design, or due to restrictions on theories to be tested. Little variation in the explanatory variables controlling for rural household characteristics may also be a source of insignificant results.

Hagen-Zanker and Siegel (2007) carried out a comparative analysis of the determinants of remittances between Albania and Moldova. They argue that the causes and patterns of migration in Albania and Moldova influence the remitting behaviour and most migrants migrate in order to remit. Nevertheless, they suggest being careful when declaring migrant's motives to remit and drawing conclusions from a few variables that can be interpreted in different ways, because a migrant may have more than one motive in mind when remitting. Their results suggest that the geographical location, economic possibilities and family situation significantly affect the place, duration and circumstances under which someone migrates and sends remittances. With regard to the characteristics of migration their results suggest that the majority of migrants are males and remittances are sent to all income groups, but in Albania, higher amounts are sent to the poorer households. Albanian migration is longer term, and higher amounts of remittances are received.

3 MODEL SPECIFICATION, THE DATA AND VARIABLE MEASUREMENT

In light of the migration and remittance theories, the empirical approach followed in this study attempts to explain the joint probability of *permanent*¹ international migration and the receipt of remittances. In order to identify the household and community characteristics which are predictive of individual migration and remittance sending, we use the probit model, where the dependent variable is a dummy variable that takes the value one if the household receives remittances from household members abroad. Specifically, the model takes form:

$$\Pr(Y=1 | X)=\Phi(X'\beta)$$

where Pr denotes probability, Φ is the Cumulative Distribution Function of the standard normal distribution, β are the parameters that will be estimated by maximum likelihood and X is a vector of explanatory variables.

¹ In this study a permanent migrant is someone who stays abroad for more than 12 months.

Data for this study come from the 2008 Albanian Living Standards Measurement Survey (LSMS), a nationally-representative survey of 3,600 households carried out by the Albanian Institute of Statistics, the World Bank and the United Nations Development Programme. The dependent variable in the Probit model is a dummy variable indicating if a household has someone permanently settled abroad and who sends remittances back home. So, the first problem lies in the identification of permanent migrants. In the migration module the head of the household is asked to list all his/her children who are not living in the household and spouse if (s)he is not living in the household. Information is also collected on when they had left the household and where they were currently living. From these two questions we derive a dummy variable taking the unit value when the household has at least one member who is actually living abroad by the time of the survey, and had not returned to live in Albania yet.

The set of variables that explain the probability to receive remittances includes individual and household characteristics and community level variables. At the household level, to account for human capital the highest level of education of the head of the household is included. The effect of education levels on the probability of migration and/or remittances may be positive or negative. Age of the head of the household is measured as a binary variable and its sign (if any), on the probability of receiving remittances remains unclear and depends on the motives behind remittance sending.

With regard to age composition shares, in both stages the number of adults from 15 to 25 and 26 to 40 years and children under 15 are included in the model. The number of adults from 15 to 40 in the household is expected to positively influence migration, while the number of children under 15 is expected to be negatively related to this. Concerning the receipt of remittances, a positive effect is expected in relation to the number of children, and a negative effect for the number of adults of working age.

The size of the household and the dependency ratio are also included in the model. The dependency ratio not only takes into account the number of children in the household, but it also accounts for the number of students and other members of the household, who do not receive income from any source. Both these variables are expected to positively influence the decision to send remittances, especially if the prevailing motive behind remittance behaviour is altruism.

The more adult household members hold a full job, the lower the probability to migrate. The expectations with regard to the direction of the relationship between this variable and the probability to receive remittances are ambiguous. A positive sign may be an indicator of self-interest motives, while a negative one provides evidence in favour of altruism. To account for this possibility a variable indicating the share of adult household members holding a full-time job is included.

Different migration and remittance patterns are also expected between rural and urban households, especially for international migration, which is argued to be a more expensive venture. This latter argument indicates higher international migration propensities for members of urban households, but no differences in the international migration patterns between

urban and rural areas may be expected if migration networks have been efficient in substantially lowering international migration costs. If this is the case, then differences in the probabilities of receiving remittances are clear reflection of different motives for sending remittances.

The welfare measure used in our study is an asset index based on a wide set of assets owned by the household and imputed by using the Principal Component Analysis (PCA). The asset index is in general utilized to classify household socioeconomic and wealth position in middle and low income countries where household income and expenditure data are invalid and unreliable, which supports its use as a proxy for household wealth in our model. The Albanian LSMS 2008 data enables the construction of an asset index using several assets on which the questionnaire contains information. This information includes data on ownership of durable and semi-durable assets, housing characteristics, and water supply system. In a second step, calculating a household assets' index involves assigning weight values to the indicator variables.

Measures of household wealth or asset ownership are sometimes considered to be potentially endogeneous in equations explaining migration decisions and remittance behaviour. To avoid possible endogeneity of the relative deprivation measure the same PCA procedure was applied to a set of assets that the households possessed in 1990. Before this year, Albania was under the communist regime and migration was an unknown phenomena. Thus, the variable could not influence the receipt or the amounts of household's actual remittances. In the same line, we construct a social capital index for the households.

A migration network proxy and a measure of relative deprivation are also included in the model. The construction of the network proxy draws heavily on the migration network theory. According to this theory, migrant networks are sets of interpersonal ties that connect migrants, former migrants, and non-migrants in origin and destination areas through ties of kinship, friendship, shared community origin. The migrant networks embody a kind of social capital that tends to lower the costs, risks and the extent of uncertainty involved in the process of international migration, increasing the likelihood of international movement (Massey et al., 1993) and even enabling the migration of the poor. As networks establish links between individuals in both the origin and destination areas, migrants are able to benefit from them in both areas. Existing migrants may provide information about available destinations, funds for travel, assistance in securing housing and employment and other fees to potential migrants. The variable included in the model is an interaction between percentage of migration population in the community and the number of household members 15-25 years old.

The relative deprivation index is constructed by using the household asset index 2008. The index was calculated at the household level by subtracting the median of primary sampling units, where 8 households were interviewed in the same period of time, from the household asset index. The median is preferred to average index in order to avoid the effect of possible outliers within the community. The constructed relative deprivation index and its square are included in the

choice equation given an expected inverse U-shape relationship with the propensity to migrate and the receipt of remittances. According to the altruism motive to remit discussed in section 2.1, households with a low negative relative deprivation index are expected to have higher propensities of remittance receipt, in order for them to improve their rating in comparison to other household members. The same may also be expected for households with high positive levels of relative deprivation, but in this case remittances are more likely to be sent for insurance and inheritance motives rather than altruism. The argument is in line with the results of Stark and Taylor (1991) who obtained an inverse 'U' shaped relation between their relative deprivation index and the probability of migration within and from Mexico to the United States.

A dummy variable indicating the presence in the household of at least one member that suffers from a chronic disease is also considered to be a potential explanatory variable. It may be argued that the presence of a member that suffers from a chronic disease in the household may decrease migration propensities of other members because of the special care needs. Finally, a dummy variable that takes the value one if the household has suffered a shock in the last 10 years is also included in the model.

4 THE REGRESSION RESULTS

An estimation of the instrumental variable probit was carried out in order to control for the possible endogeneity of the welfare proxy, the household asset index 2008. The results of this model indicate that this variable is not causing endogeneity, so a probit estimation is used instead. Table 1 shows the results of the estimation of the Probit model that explains the probability of international migration and the receipt of remittances. Household welfare is theoretically related to self-interest and the bequest motive for sending remittances, which are presumed to motivate migrants to remit for inheritance. So, migrants with a bequest motive should be more likely to send remittances, and even send more if they have wealthier parents. Some studies suggest investigating the possibility of a non-linear relationship between migration and welfare measures arguing that the poorest of the households are too poor to migrate because they cannot afford the costs of migration, while the richest have no incentives (Lucas, 2005). Results suggest that welfare variables may have positive (Hodinnot, 1994; de la Briere, 2002; Pleitez-Chavez, 2004; Schrieder and Knerr, 2000; Hagen-Zanker and Siegel, 2007), or negative effects (Agarwal and Horowitz, 2002; Durand et al,

1996; Osaki, 2003), while others conclude that the effects are not significant (Osaki, 2003; Holst and Schrooten, 2006).

Our results provide evidence on the insignificant relationship between these variables. We included a measure of household assets ownership and per capita consumption levels and tested for a non-linear quadratic relationship, but the results suggested that the linear and squared term of the household asset index and of the per capita consumption levels were jointly insignificant and consequently they were excluded from the model. It is worth mentioning though that the effect of the household wealth is difficult to disentangle when part of household's assets is a consequence of past remittances. In this case it is unclear if including a measure of household's wealth controls for effects of wealth alone, or past migration effects as well. For this reason, we explicitly include controls for past migration in our model, in order to have unbiased coefficients on the wealth variables. The variable on the previous migration experience has the expected positive sign and is statistically significant.

The results also indicate that the variables on the characteristics of the head of the household are highly significant, except for the male-headed households. Experience is considered a key determinant of earnings in human capital models (Sjaastad, 1962; Mincer, 1974). In the absence of data on the migration duration of the individual migrants, in household level studies experience is usually proxied by the age of the household head. It may also be argued that households with older heads are likely to produce more migrants because they have more household members between 15 and 30 years old, which may be considered as the prime age span of migration. Empirical results suggest that the age of the head of household has the expected positive sign (Adams, 2004, 2006; Osili 2007; Adams et al., 2008), although it does not always influence the decision to migrate (Osili, 2007; Adams et al., 2008). The age of the head of the household is also expected to be related to remittance receipt and the amount received. More evidence in favour of the altruism motive for remitting is consistent with households with older heads receiving more remittances. The empirical results indicate that this variable does not affect the decision to remit (de la Briere et al., 1997, 2002; Agarwal and Horowitz, 2002; Pleitez-Chavez, 2004; Pfau and Giang, 2010), which is more consistent with the investment motive for sending remittances.

Table 1. Estimation of the probability to migrate and receive remittances

	dF/dx*	Std. Err.	P> z
Characteristics of the head of household			
Head of household is married*	0.101	0.027	0.003
Male-headed*	0.053	0.048	0.277
Head of household under 65*	-0.097	0.036	0.004
Head of household has 8 years school	-0.092	0.025	0.000
Head of household has a vocational	-0.120	0.022	0.000
Head of household has secondary	-0.129	0.020	0.000
Head of household has university of post-	-0.120	0.021	0.000
Household characteristics			
Size of the household	-0.094	0.011	0.000
Number of members 15-25 yrs	0.097	0.038	0.010
Square of number of members 15-25 yrs	-0.014	0.016	0.357
Number of members 26-40 yrs	0.047	0.014	0.001
Square of number of members 26-40 yrs	0.001	0.000	0.038
Number of children under 15 yrs	-0.056	0.020	0.004
Square of number of children under 15	0.020	0.005	0.000
Share of members in full-time	-0.238	0.184	0.193
Gender ratio	0.096	0.046	0.036
Dependency ratio	-0.069	0.038	0.067
At least a member with a chronic	0.039	0.019	0.035
Social capital index	-0.007	0.006	0.232
Square of social capital index	0.000	0.002	0.836
Suffered a shock in the last 10 years*	0.012	0.017	0.467
Previous migration experience*	0.060	0.020	0.004
Relative deprivation Index	0.026	0.007	0.000
Square of relative deprivation index	-0.004	0.001	0.003
Migration percentage*Members 15-25	0.000	0.000	0.818
Location			
Urban*	-0.087	0.021	0.000
Coastal*	0.102	0.033	0.001
Central*	0.021	0.032	0.496
Mountain*	-0.060	0.031	0.081
Observations	=	3,599	
Wald chi2(29)	=	437.37	
Prob > chi2	=	0.0000	
Pseudo R2	=	0.1904	
Log pseudolikelihood	=	-1,275,582	

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

In difference from these studies and despite the non-linearities in the effects of age reflected through the non-linear functional form of the empirical model, we include certain age brackets rather than a continuous variable for the age of the head of the household. Our results indicate that households whose heads are married and over 65 years old have higher international migration and remittance

propensities. This may be related to the fact that these households may have more adult members at the prime age span of migration and supports the altruism motive behind remittances.

Human capital variables are likely to have positive effects on migration if there are higher possibilities of employment and expected income-earning in destination areas compared to the origin. Considering the effect of the household head's education on migration, on the one hand, more educated parents or those who own a business may encourage their children to study more and seek opportunities in the country or contribute in their business. On the

other hand, relatively low returns to education in the origin compared to destination countries may increase migration propensities. Most of the empirical studies indicate a statistically significant impact of education on migration, although they provide conflicting findings with regard to the sign of the education variable. Some research suggests that migration is negatively associated with education (Mora and Taylor, 2006; Boucher et al., 2005; Borjas, 1990), while others suggest it being positively associated (Kanaiaupuni, 2000; Garip 2006; Palloni et al., 2007; Zhu and Luo, 2008). The results in our study are in line with those of the first group of these studies. We conclude that in Albania migration and remittances are less likely to happen to households with more educated heads in comparison to households whose heads have few years or no education at all, indicating that migration and remittances negatively select on education of the household heads. Similar results are found for Albania (de Coulon and Piracha, 2005; Germeji and Swinnen, 2005; Piracha and Vadean, 2010). The findings suggest that Albania is not facing a brain drain problem, which is common for countries with high migration flows. But this finding has to be taken with caution, as a household's educational attainment has been proxied by the education level of the head of the household due to lack of data.

Household demographic characteristics are also hypothesized to affect the probability of migration and the receipt of remittances. The household size (Gubhaju and de Jong, 2009; Phuong et al., 2008), and the age composition shares are usually included in the model with the expectation that households with many young adults are more likely to send someone abroad because of surplus labour (Phuong et al., 2008). The number of children of different age groups and the dependency ratio are among the most commonly used variables (Katz, 2000; Garip, 2006; del Rey Poveda, 2007; Acosta et al., 2007; Zhu and Luo, 2008; Rainer and Siedler, 2008). The number of children is expected to negatively affect the migration decision, especially in small size households. If there are more adult females in the household then the propensities to migrate increase as only a few of them usually care for the children or elderly. To control for this some studies include the gender ratio (Katz, 2000; Garip, 2006). The household size and/or the dependency ratio may be also related with remittances sent for altruism. More members in the household and especially more children may mean that the migrant feels responsible for their wellbeing and thus remits more. Most authors find a positive effect of the household size in estimations of the probability and level of remittances (Itzigsohn, 1995; Osili, 2007), and a negative effect of the dependency ratio as expected (Agarwal and Horowitz, 2002; Osaki, 2003). Others report insignificant effects of these variables (Durand et al., 1996; Sela, 2004; Craciun, 2006; Hagen-Zanker and Siegel, 2007).

The results in table 1 indicate that the size of the household has an unexpected negative effect on the probability to migrate and receive remittances. The number of adult members aged 15-25 years has a U-shaped relationship and the number of adults 26 to 40 has a positive nonlinear relationship as expected. The number of children under 15 years also has a significant U-shaped relationship with migration and remittance receipt. The gender ratio has a positive and significant relationship with migration and remittances, while the dependency ratio which accounts for the presence of children, students and adults who do not receive any income is negative, but insignificant.

Several measures of employment are also included in models of migration propensities. Some of these measures include the share of household members working in wage employment (Phuong et al., 2008; Zhu and Luo, 2008), or occupational status (Rainer and Siedler, 2008; Pfau and Giang, 2010). Estimates suggest that these variables have negative effects on migration as expected. With regard to the determinants of remittances, Pfau and Giang (2010) control for the head of the household's employment status and the results suggest that the head of households tend not to work when they receive international remittances. They argue that either they become lazy and less likely to work, or they are unable to work and for this reason they receive more remittances. The results in table 1 indicate that the share of adults holding a full time job has the expected negative sign, but its effect is insignificant on the probability to migrate and receive remittances.

The explicit variables used to measure the importance of social capital as embedded in the theory of migration networks (Massey et al., 1993) vary between among studies. Such diversity may originate from the data availability, different cultures, contexts, and models, as well as from the broadness of the concept itself. To account for migration networks del Rey Poveda (2007), Richter and Taylor (2007), and Palloni et al. (2007) include a variable indicating the household's history of migration. Garip (2006) uses the percentage of community's households receiving remittances, and the frequency of visits paid by the migrants in the last 10 years, which is also expected to affect remittances. Less frequent visits may weaken the ties to the home country and household, lowering the importance of altruism (Niimi et al., 2008). We control for the effect of two kinds of social capital, in the home and in the host country. The impact of the social capital owned by the household in the home country and the impact of the proxy of migration networks are not statistically significant, i.e. in Albania the social capital does not influence international migration or the receipt of remittances.

When studying the determinants of remittances many researchers have in focus the co-insurance motive. This is measured by including a household shock in the model, such as illness, or number of lost working days. For example, de la Briere et al. (1997, 2002) conclude that they lead to a lower probability of migration and higher probability of remittances. Considering the impact of a shock suffered by the household in the last ten years, a dummy variable constructed accounting for different types of shock, the results suggest that in Albania suffering a shock has a statistically insignificant impact on the propensities to migrate and receive remittances. The presence of at least one household member suffering from a chronic disease is also included in the model and the results suggest that this variable positively influences the probability to send someone abroad and receive remittances which supports the co-insurance motive behind migration.

This study proposed the inclusion of a relative deprivation index and theoretically argued about its non-linear relationship with migration and remittances. According to the altruism motive to remit discussed earlier, households with a low negative relative deprivation index are expected to have higher propensities of migration and remittance receipt, in order for them to improve their rating in comparison to other household members. The same may also be expected for households with high positive levels of relative deprivation, but in this case remittances are more likely to be sent for insurance and inheritance motives rather than altruism. The em-

pirical results indicate that it has the expected significant inverted U-shaped relationship with migration and remittances.

Finally, dummy variables for urban area and regional dummies are included to control for other geographical differences affecting the incentive to migrate. The geographical location of receivers can also account for part of the variation in remittance patterns that is unexplained by household or individual factors, because it may act as a proxy for other socio-economic factors and norms at the community level. We control for the effect of the rural/urban location and three regional dummies to estimate the specific effect of the receiving community's development level. The results indicate that households in the urban areas have significantly lower migration and remittance receipt propensities in comparison to the rural households. Furthermore, compared to Tirana, migration and remittance propensities are higher among households in the Coastal region, but there is no difference in the migration and remittance propensities with the Mountain and Central regions. These results are in line with those found in other studies (Funkhouser, 1995; Hagen-Zanker and Siegel, 2007; Lerch and Wanner, 2006; Niimi et al., 2008; Pfau and Giang, 2010).

5 CONCLUDING REMARKS

This analysis has examined the determinants of migration and remittances at the household level in Albania using a representative survey of 3,599 households. The relationship between household income and migration is insignificant. The results of the probit regression including per capita consumption and a measure of household asset index to control for household wealth indicate that the relationship is insignificant. These results hold even when controlling for non-linear quadratic relationship between these variables and the propensities to migrate and receive remittances. In difference from other studies, we provide no evidence of the "migration hump".

The age and marital status of the head of the household are significant, while gender is not significant. The human capital variables have significant effects, showing that households with more educated heads have lower migration and remittance propensities. This result indicates that Albania is not facing a brain drain problem, but this finding has to be taken with caution, because the highest education level of the head of the household is a proxy for the migrant's education level. This result is in line with the ones found in other studies of migration determinants in Albania (de Coulon and Piracha, 2005; Germenji and Swinnen, 2005; Piracha and Vadean, 2010).

The household characteristics, such as size, number of adults 15 to 40 and number of children are statistically significant, supporting the household approach. The results also suggest that the gender ratio has a positive and significant relationship with migration and remittances. Kotorri (2010) controls for a different gender ratio and finds a negative influence of the gender composition on the household migration behaviour in Kosovo. The presence of at least one household member suffering from a chronic disease has important positive influence on the probability to send someone abroad and receive remittances, which supports the co-insurance motive behind migration. A shock suffered by the household has a

statistically insignificant impact on the propensities to migrate and receive remittances.

The share of adults holding a full-time job is not significant. We also attempted to include other employment variables in the model, but the results were consistent among specifications, indicating that the labour supply surplus has no effect on international migration and the receipt of remittances. This result is different from those of other studies reviewed here. Variables introduced to capture the effect of social capita on migration do not show any significant influence.

The majority of the studies on the determinants of migration and remittances do not include any measure of relative deprivation. The relative deprivation plays an important role in determining migration and remittances. The constructed relative deprivation index and its square are included in the equation given an expected inverse U-shape relationship with the propensity to migrate and the receipt of remittances. The results obtained confirm the expectations and are in line with the results of Stark and Taylor (1991) who obtained an inverse 'U' shaped relation between their relative deprivation index and the probability of migration within and from Mexico to the United States. When modelling temporary and permanent migration in Albania, Carletto et al. (2004) control for the effect of relative deprivation by using a different index, but their results are different.

As expected, households living in urban areas have a lower probability of migration and remittance receipt. Other studies also control for regional characteristics and find support in favour of their importance. Compared to Tirana, households in the Central region have higher probability of migration and remittances.

REFERENCES

- [1] Adams, Jr., Richard H., Cuecuecha, A., and Page, J., "Remittances, Consumption and Investment in Ghana." World Bank Policy Research Working Paper Series, 2008. Available at SSRN: <http://ssrn.com/abstract=1093127>
- [2] Adams, R., "Remittances and Poverty in Guatemala". World Bank Policy Research Working Paper No. 3418, 2004.
- [3] Adams, R., "Remittances and Poverty in Ghana". World Bank Policy Research Working Paper, No. 3838, 2006.
- [4] Agarwal, R., and Horowitz A.W., "Are international remittances altruism or insurance? evidence from Guyana using multiple-migrant households." World Development. vol. 30, no. 11, pp. 2033-2044, 2002.
- [5] Borjas, G. J., *Friends or strangers: The impact of Immigration on the U.S. Economy*. N.Y.: Basic Books Inc., 1990.
- [6] Boucher, S., Stark, O., and Taylor, J. E., "A Gain with a Drain? Evidence from Rural Mexico on the New Economics of the Brain Drain." Paper prepared for the 14th World Congress of the International Economic Association, Marrakech, Morocco, August 29-September 2, 2005.
- [7] Carletto, C., Davis, B., Stampini, M., and Zezza, A., "A Country on the Move: Internal Mobility and International Migration in Post-Communist Albania" FAO (mimeo), 2004. Rome (Italy).
- [8] Craciun, C., "Migration and Remittances in the Republic of Moldova: Empirical Evidence at Micro Level". National University "Kyiv-Mohyla Academy", 2006.
- [9] de Coulon A., and M. Piracha., "Self-Selection and the Performance of Return Migrants: The Source Country Perspective." Journal of Population Economics, vol. 18, no. 4, pp. : 779-807, 2005.
- [10] de la Briere, B., Sadoulet, E., de Janvry, A., and Lambert, S., "The roles of destination, gender, and household composition in explaining remittances: An analysis for the Dominican Sierra". Journal of Development Economics, vol. 68, no. 2, pp. 309-28, 2002.
- [11] de la Briere, B., Janvry, A. d., Lambert, S., and Sadoulet, E., "Why do migrants remit? An analysis for the Dominican Sierra." FCND Discussion Paper, No. 37, 1997.

- [12] Durand, J., Kandel, W., Parrado, E. A., and Massey, D. S., "International migration and development in Mexican communities." *Demography*, vol. 33, no. 2, pp. 249-264, 1996.
- [13] Funkhouser, E., "Remittances from International Migration: A Comparison of El Salvador and Nicaragua." *The Review of Economics and Statistics*, vol. 77, no. 1, 1995.
- [14] Garip, F., "Social and Economic Determinants of Migration and Remittances: An Analysis of 22 Thai Villages." Princeton University Press. Mimeo, 2006.
- [15] Germenji, E., and Swinnen, J., "Human Capital, Market Imperfections, Poverty and Migration: Evidence from Albania." Lycos Discussion Paper No. 157, Katholieke Universiteit, Leuven, 2005.
- [16] Germenji, E., Beka, I., and Sarris, A., "Estimating remittance functions for rural-based Albanian emigrants." Working paper, ACE Research Project, P97-8158-R. 2001.
- [17] Gubhaju, B., and de Jong G. F., "Individual vs. Household Decision Rules: Gender and Marital Status Differences in the Intention to Move in South Africa." *International Migration* vol. 47, no. 1, pp. 31-61, 2009.
- [18] Hagen-Zanker J., and Siegel, M., "The determinants of remittances: A comparison between Albania and Moldova." MGSOG Working Paper 2007-007, 2007.
- [19] Hoddinott, J., "A model of migration and remittances applied to western Kenya." *Oxford Economic Papers*, vol. 46, no. 3, pp. 459-476, 1994.
- [20] Holst, E., and Schrooten, M., "Migration and money- What determines remittances? Evidence from Germany." Institute of Economic Research, Hitotsubashi University, Discussion Paper Series, 477, 2006.
- [21] Itzighson, I., "Migrant Remittances, Labour Markets, and Household Strategies: A Comparative Analysis of Low Income Household Strategies in the Caribbean Basin." *Social Forces*, vol. 72, no. 2, 1995.
- [22] Kanaiaupuni, S. M., "Reframing the Migration Question: An Analysis of Men, Women, and Gender in Mexico." *Social Forces*, vol. 78, no. 4, pp. 1311-47, 2000.
- [23] Katz, E.G., "Individual, Household and Community-Level Determinants of Migration in Ecuador: Are there Gender Differences?" Paper presented at the Annual Meeting of the Population Association of America. March 23-25, Los Angeles, CA, 2000.
- [24] Konica, N., and Filer R. K., "Albanian Emigration: Causes and Consequences." *South-Eastern Europe Journal of Economics*, vol. 1, pp. 75-98, 2009.
- [25] Kotori, M., "The Probability of Economic Emigration in Kosovo." *Eastern European Economics*, vol. 48, no. 6, November-December 2010, pp. 17-35, 2010.
- [26] Lerch, M., and Wanner, P., "Les Transferts de Fonds des Migrants Albanais," *Etudes du SFM*, 45, Neuchatel, Swiss Forum for Migration and Population Studies, 2006.
- [27] Lianos, T. P., & Cavounidis, J., "Immigrant remittances, stability of employment and relative deprivation..." Paper presented at the New Perspectives on Albanian Migration and Development, Korçë, Albania. 2004.
- [28] Massey, D.S., Arango, J., Hugo, G., Kouaouchi, A., Pellegrino, A., and Taylor, J.E., *Theories of International Migration: A Review and Appraisal*. Population and Development Review, vol.19, no. 3, pp:431-66, 1993.
- [29] Mincer, J., *Education and Earnings*. Columbia University. 1974.
- [30] Mora, J., and Taylor, J. E., "Determinants of Migration, Destination, and Sector Choice Disentangling Individual, Household, and Community Effects." In *International Migration, Remittances, and the Brain Drain*, ed. Gaglar Ozden and Maurice Schiff. New York: Palgrave Macmillan, 2006.
- [31] Niimi, Y., Pham, T. H., and Reilly, B., "Determinants of Remittances: Evidence from Internal Migrants in Vietnam", *Asian Economic Journal*, vol. 23, no. 1, pp. 19-39, 2008.
- [32] Osaki, K., "Migrant remittances in Thailand: Economic necessity or social norm." *Journal of Population Research*, vol. 20, no. 2, pp. 203-222, 2003.
- [33] Osili, U. O., "Remittances and savings from international migration: Theory and evidence using a matched sample." *Journal of Development Economics*, vol. 83, 446-456, 2007.
- [34] Palloni, A., Massey, D. S., Ceballos, M., Espinosa, K., Spittel, M., "Social capital and international migration: A Test Using Information on Family Networks." *The American Journal of Sociology*, vol. 106, no. 5, pp. 1262-1298., 2007.
- [35] Pfau, W. D. and Giang L. T., "Determinants and impacts of international remittances on household welfare in Vietnam", *International Social Science Journal*, Vol.60, no. 197-198, pp. 431-443, September/December, 2009.
- [36] Phuong, Nguyen Thu, Tran N. T. M. Tam, Nguyen and Nguyet T., "Determinants and Impacts of Migration in Vietnam." DEPOCEN Working Paper Series No. 2008/01, 2008.
- [37] Piracha, M., and F. Vadean, "Return Migration and Occupational Choice: Evidence from Albania". *World Development*. vol. 38, no. 8, pp. 1141-1155, 2010.
- [38] Pleitez-Chavez, R.A., "Remittances as a Strategy to Cope with Systemic Risk: Panel Results from Rural Households in El Salvador." Unpublished PhD, Ohio State University, Columbus. 2004.
- [39] Rainer, H. and Siedler T., "Social Networks in Determining Migration and Labor Market Outcomes: Evidence from the German Reunification," *Economics of Transition*, vol. 17, no. 4, pp. 739-767, 2009.
- [40] Schrieder, G., and Knerr, B., "Labour Migration as a Social Security Mechanism for Smallholder Households in Sub-Saharan Africa: The Case of Cameroon." *Oxford Development Studies*, vol.28, no. 2, pp. 223-236, 2000.
- [41] Sela, "Global Recession, Migration and Remittances: Effects on Latin American and Caribbean Economies." *Sistema Economico Latinoamericano y del Caribe*, 2004.
- [42] Sjaastad, L. A. "The Costs and Returns of Human Migration." *Journal of Political Economy*, vol. 70, no. 5, pp. 80-93, 1962.
- [43] Stark O., and Taylor, J.E., "Migration incentives, migration types: The role of relative deprivation", *The Economic Journal*, vol. 101, no. 408, pp. 1163-78, 1991.
- [44] Zhu, N., and Luo, X., "The Impact of Remittances on Rural Poverty and Inequality in China." *World Bank Policy Research Working Paper No. 4637*, 2008.

Hydraulic Regenerative Braking System

ER. Amitesh Kumar

ABSTRACT- Hydraulic regenerative braking system is an important branch of hybrid technology, which has the advantage of high power density and the ability to accept the high rates/high frequencies of charging and discharging, therefore hydraulic regenerative braking technology is well suited for off-road vehicles and heavy-duty trucks. Relatively lower energy density and complicated coordinating operation between two power sources require a special energy control strategy to maximize the fuel saving potential. This paper presents a new configuration of parallel hydraulic regenerative vehicle (PHRBV) to improve the braking energy regenerated potential and engine work efficiency. Based on the analysis of optimal energy distribution for the proposed PHRBV over a representative urban driving cycle, a fuzzy torque control strategy based on the vehicle load changes is developed to real-time control the energy distribution for the proposed PHRBV. Simulation results demonstrate that the proposed PHRBV with torque control strategy takes advantage of the high power density and efficiency characteristics of the hydraulic regenerative braking system minimizes the disadvantages of low energy density and effectively improves the fuel economy of PHRBV.

Keywords – Fuzzy logic controller System, Fuzzy control principle, Hydraulic system schematic, parallel hydraulic regenerative braking system

INTRODUCTION

ONE must first understand what a Hydraulic regenerative braking system is, and how it works. The basic idea of a Hydraulic regenerative braking system is that when the vehicle slows down or decelerates, it will store the kinetic energy that was originally momentum as potential energy in the form of pressure. This is done by using a displacement pump to pump hydraulic fluid into an accumulator. When the vehicle accelerates, the pressure is released from the accumulator which will spin the drive shaft and accelerate the vehicle. Thus the engine remains idle while the pressure is released and when the accumulator is empty, or the desired speed is achieved, the engine will then engage in order to maintain a constant velocity, or to accelerate the vehicle beyond what the capacity of the accumulator was capable of. fig 1.1 shows the basic construction of it.

- Er.Amitesh Kumar(Mechanical Engineer) Masters of Engineering in Automotuve Engineering from Sinhgad Academy of Engineering, Pune, India.
E-mail: amitesh.parasar@gmail.com

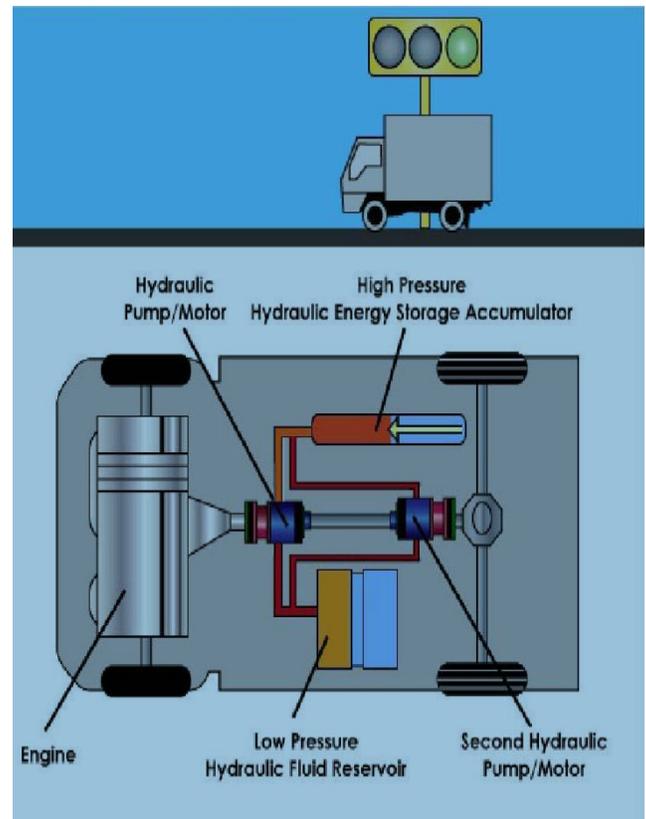


Figure 1.1: hydraulic regenerative braking system

1.1 Problem Statement

Inefficient fuel usage is often unavoidable for many vehicles. Most specifically, vehicles that operate under frequent stop and go conditions, such as delivery

vehicles, are most affected by these inefficiencies. With increasing fuel prices and inefficient fuel use, there is an obvious need for a more resourceful solution. The solution is Hydraulic regenerative braking system.

1.2 Objectives

The primary objective of this topic is to validate that a Hydraulic regenerative braking system can increase the stop and go fuel efficiency of a vehicle by 32%.

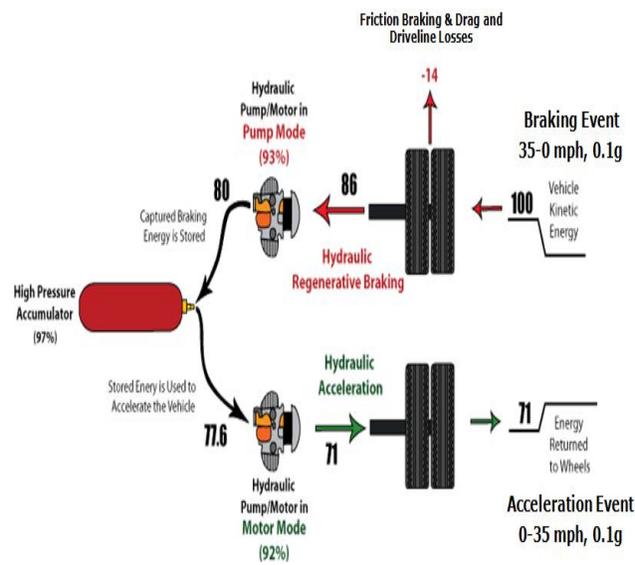


Figure 1.2: Efficiencies while braking/ accelerating

2. DESIGN FUNCTIONALITY

The design is transparent so that the control of this system functions as intuitively as possible. It was designed to be controlled by gas and brake pedals so that any new user will be familiar with its control. The design of the hydraulic system was created so it in no way hinders the performance or integrity of the vehicle. Not only will the system be predictable in use, it will be reliable and consistent.

2.1 Scope of the topic

The scope of the topic was to test hydraulic regenerative braking technology, rather than designing a marketable system. Much of the time was spent on better understanding and implementing the knowledge and

design aspect of the hydraulic regenerative braking system.

2.2 Hybrid Background

2.2.1 Types of Hybrids

Hybrid vehicles are becoming more and more common in the auto industry. A hybrid vehicle is simply a vehicle that operates using a primary engine and secondary energy storage device. While electric hybrid vehicles are the most familiar and have been commercially produced, Hydraulic regenerative braking system technology is being investigated as a better hybrid option. Hydraulic regenerative braking vehicles are being introduced into the industry by companies such as Parker Hannifin Corporation and Eaton Corporation. Hydraulic regenerative braking system can be divided into two different system types: parallel and series.

In a parallel hydraulic hybrid, the hydraulic components are connected to the conventional transmission and driveshaft. While the engine is always running and consuming fuel, the hydraulic system is able to assist in acceleration when the engine is working its hardest, thus increasing fuel efficiency. Series hydraulic hybrid systems use the same basic principles as parallel systems, but do not utilize the conventional transmission or driveshaft. The hydraulic system transmits all power directly to the wheels.

2.2.2 Benefits of Hydraulic Breaking System

2.2.2.1 Less Energy Conversions

A hydraulic hybrid offers a number of advantages over electric hybrids. The largest of these advantages is that the system employs less energy conversions, leading to higher efficiencies. An electric hybrid converts kinetic energy to electrical energy and then to chemical energy that is stored in a battery. A hydraulic hybrid converts kinetic energy to fluid pressure energy and this is how the energy is stored. This results in one less energy form conversion to help the hydraulic regenerative braking system achieve higher efficiency than other types of hybrids.

2.2.2.2 Regenerative Braking

Hydraulic Regenerative braking is a large advantage to a hybrid system, especially when the vehicle is subject to frequent stops. Normally, in a conventional vehicle, all of the kinetic energy is lost to heat, an irreversible process. A hybrid however captures some of this energy through regenerative braking to be reused the next time the vehicle accelerates. Through simulations, it was estimated that 70% of the kinetic energy of the vehicle would be recycled through a regenerative braking event with the hydraulic regenerative braking system.

2.2.2.3 Optimal Engine Speed

Operating the engine of a hydraulic regenerative braking vehicle at its optimal operating point is of value because the engine can generate the most power per flow rate of fuel. To find this point for the engine used in this project, the following data was obtained from the manufacturer: engine power versus engine speed, engine torque versus engine speed, and engine fuel consumption rate versus engine speed. These data graphs are presented below in Figure 2.1, Figure 2.2, and Figure 2.3. The engine used in this project is represented by the orange line in all three figures. From this, a ratio was calculated at each engine speed of engine power or torque per volumetric flow rate of fuel. This was then graphed to find the optimum operation point. From Figure 2.4 and Figure 2.5, it was determined that the slower the engine runs, the better the torque per fuel consumption rate. However, the power graph shows a speed at which the engine maximized the amount of power obtained per fuel flow rate. From this data, the engine would be tuned to run at about 2200 rpm when pumping hydraulic fluid. This speed is close to the optimal operating speed for our engine, thus providing the best use of the fuel consumed by the engine.

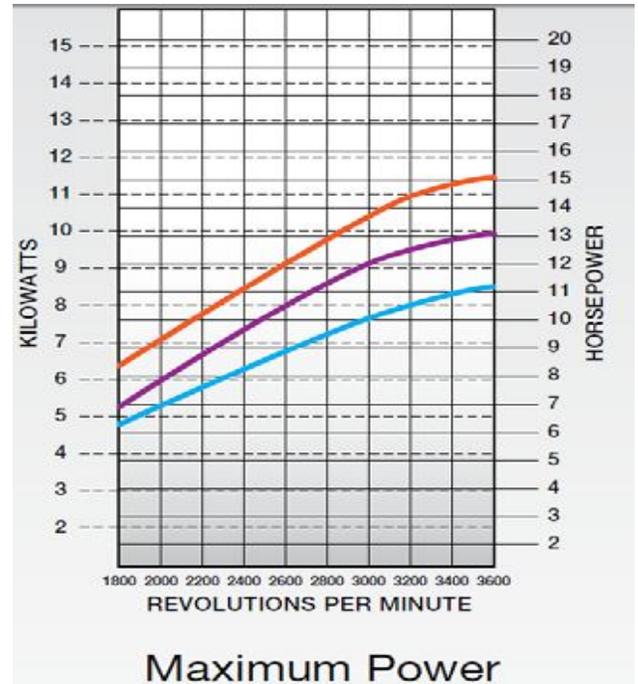


Figure 2.1: Engine power data

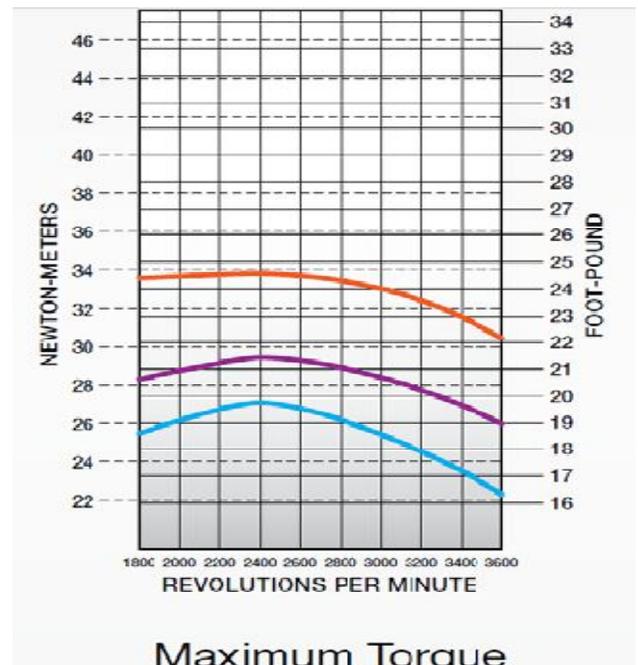


Figure 2.2: Engine torque data

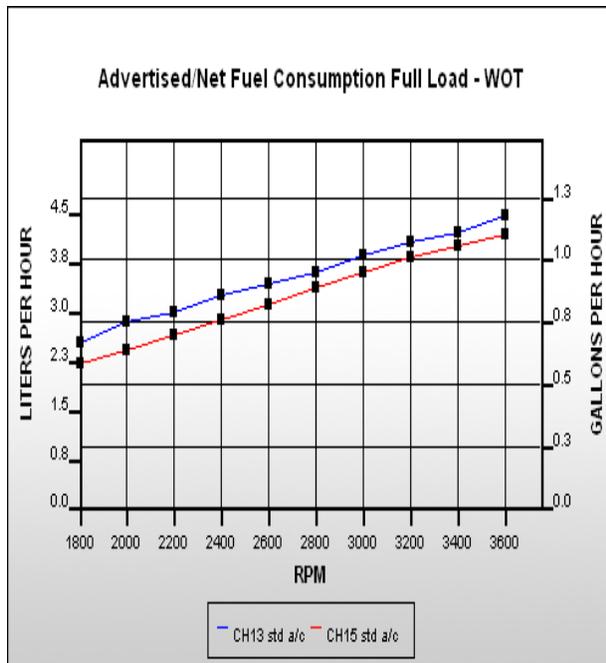


Figure 2.3: Engine fuel consumption data

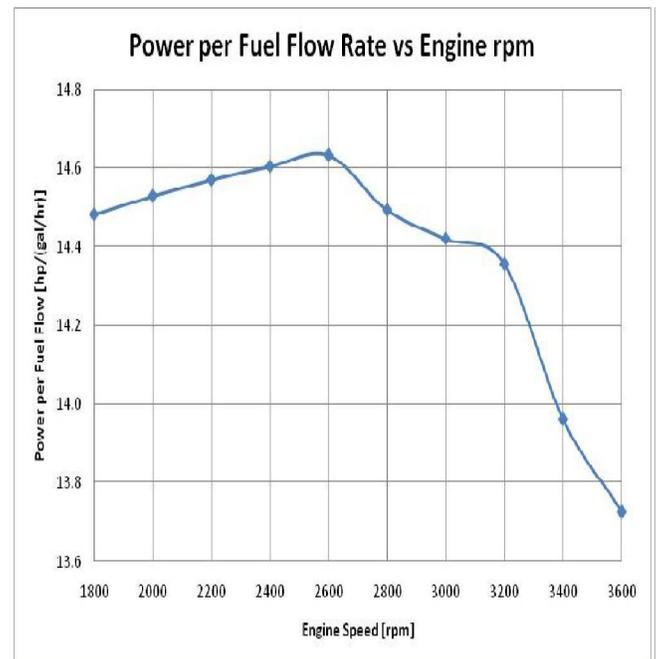
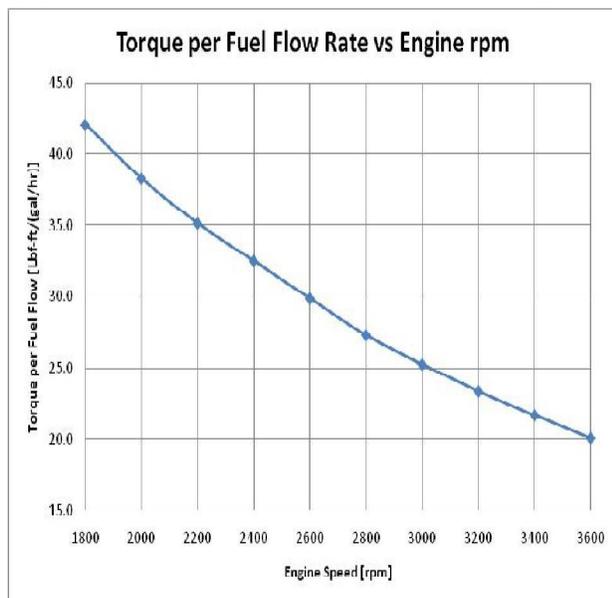


Figure 2.5: Graph of power per fuel flow rate versus engine speed

Figure 2.4: Graph of torque per fuel flow rate versus engine



speed

2.3 Hydraulic regenerative braking system

A Hydraulic regenerative braking system schematic was created for a series hydraulic hybrid vehicle. This schematic was designed to include acceleration and braking control using hydraulic flow control valves, regenerative braking using check valves, and forewords and reverse directions using a directional selector valve. An important safety feature of the hydraulic schematic is the high pressure relief valve, which ensures that the pressure in the system never reaches an unsafe level. A copy of this schematic can be found in Figure 2.6

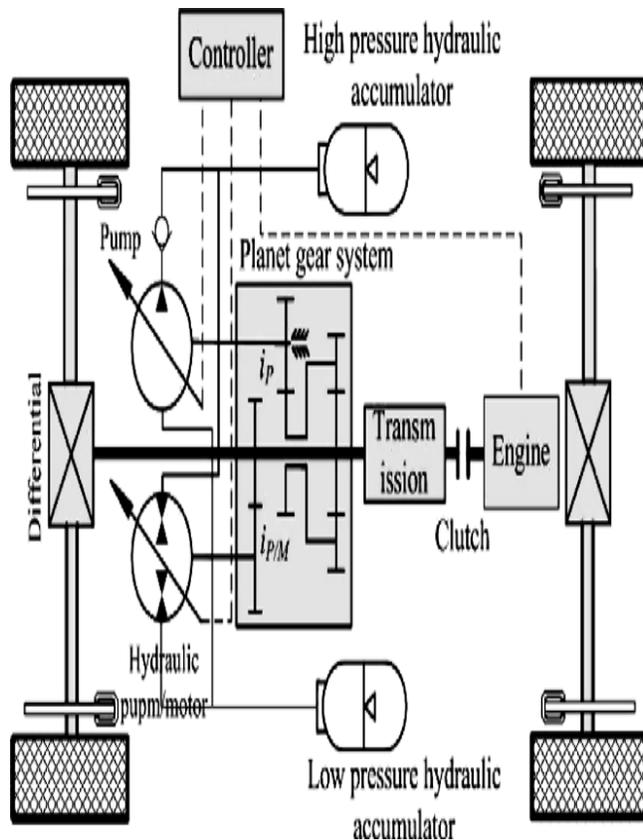


Fig. 2.6 New configuration of parallel hydraulic regenerative braking system.

Fig.2.6 presents a new type of configuration for hydraulic regenerative braking system which consists primarily of an engine, a high-pressure accumulator, a low pressure reservoir, a variable displacement hydraulic pump, and a variable displacement hydraulic pump/motor unit, clutch, transmission and differential. The hydraulic pump/motor and hydraulic pump are coupled to the propeller shaft via a planetary gear system. During deceleration, the hydraulic pump/motor decelerates the vehicle while operating as a pump to capture the energy normally lost to friction brakes in a conventional vehicle. Also, when the vehicle brake is applied, the hydraulic pump/motor uses the braking energy to charge the hydraulic fluid from a low pressure hydraulic accumulator into a high-pressure accumulator, increasing the pressure of the nitrogen gas in the high-pressure accumulator. The high pressure hydraulic fluid is used by the hydraulic pump/motor unit to generate torque during the next vehicle acceleration. It is designed and sized to capture braking energy from normal, moderate braking events and is supplemented by friction brakes for aggressive braking. Cruise conditions, the hydraulic pump works for charging the hydraulic accumulator,

meantime, adjust the engine working point onto the optimal fuel consumption region. When the hydraulic accumulator pressure reaches the highest initiative charging pressure value, the vehicle is driven by the hydraulic accumulator and hydraulic pump/motor. The introduction of hydraulic pump minimizes the lower energy density disadvantage of the accumulator and makes the engine work in high efficiency region through the initiative charging function.

3.1. Impact of pump/motor install position on braking energy regeneration

Allocate the hydraulic pump/motor behind the front differential (the front-wheel-drive model) has a greater potential in braking energy regeneration because of the increasing of the front axle load during braking. The front wheel braking force F_{f1} and the rear wheel braking force F_{f2} are given by [Eq. (1) and (2)], respectively.

$$F_{f1} = \varphi \left(Gb + m \frac{du}{dt} h_g \right) / L$$

$$F_{f2} = \varphi \left(Ga - m \frac{du}{dt} h_g \right) / L$$

where φ is the friction coefficient between tire and road surface, a is the distance from vehicle center of gravity to front axle center line, b is the distance from vehicle center of gravity to rear axle center line, L is the wheel base and h_g is the height of the center of gravity. During the course of braking, the front axle load increases and the rear axle load decreases, therefore, install pump/motor behind the front differential has greater potential in braking energy regeneration. The regenerative braking torque distributions under different pump/motor install position are shown in Figs. 3.1 and 3.2 the impact of hydraulic pump/motor install position on braking energy regeneration under different driving cycles is shown in Fig.3.3

3.2. Operating conditions of PHRBV

Parallel hydraulic hybrid vehicle mainly includes the hydraulic driving condition, cruise condition, accelerating/climbing condition and regenerative braking condition.

3.2.1. Hydraulic drive condition

When vehicle starts and the pressure of hydraulic accumulator is higher than the minimum working pressure value, the commanded power of vehicle is supplied only by hydraulic pump/motor. When the pressure in hydraulic accumulator declines to the minimum working pressure value, the vehicle is switched to cruise condition.

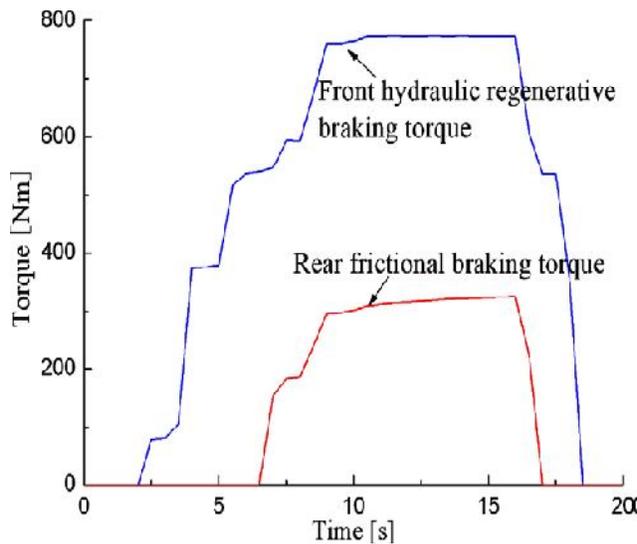


Fig. 3.1. Regenerative braking torque distribution of front-wheel-drive model.

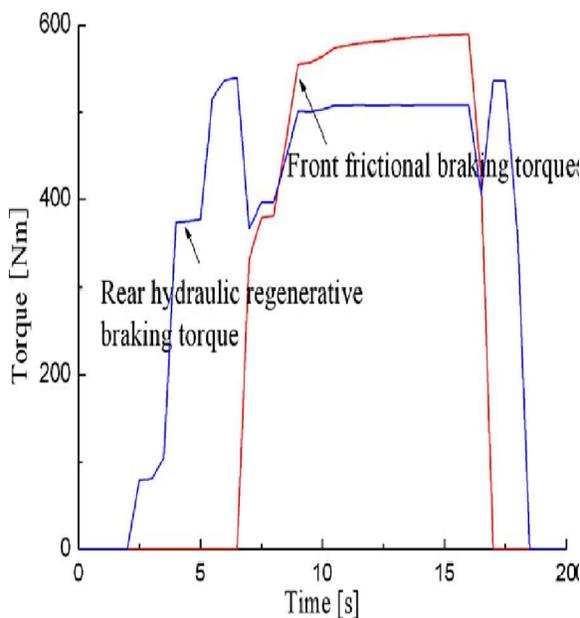


Fig. 3.2 Regenerative braking torque distribution of rear wheel drive.

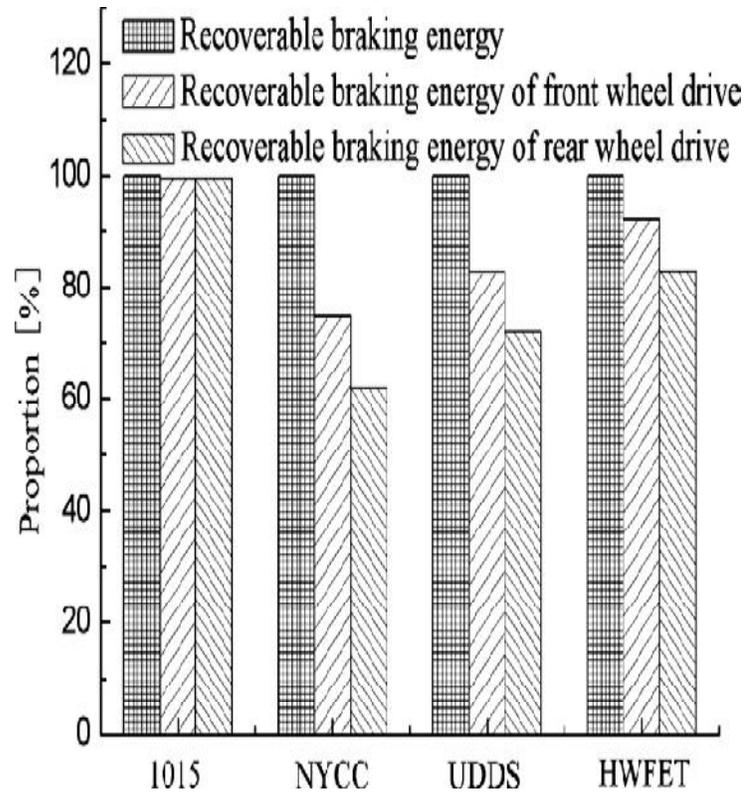


Fig. 3.3 Impact of hydraulic pump/motor configuration on braking energy regeneration.

3.2.2. Low-speed cruise condition

During the course of low-speed cruise condition, major output power of the engine is used for driving vehicle, the others power is used for charging the hydraulic accumulator, meantime, adjusts the engine load. The torque relationship is shown as [Eq. (3)]

$$T_d = T_{eng} i_g \eta_e \eta_t - T_P i_P \eta_P - T_{d,l}$$

where T_d , T_{eng} and T_P are the torque of differential, the engine output torque and hydraulic pump charging torque, respectively, i_g is the velocity ratio of transmission, i_P is the velocity ratio of hydraulic pump to the propeller shaft, η_P is the efficiency of hydraulic pump to the propeller shaft, η_e , η_t are the efficiency of engine and the efficiency of transmission and $T_{d,l}$ is the torque loss due to friction and churning loss for the transmissions to differential.

3.2.3. Middle and high-speed cruise condition

During the course of middle-speed and high-speed cruise course, in order to enable the engine work at a relatively middle speed region, the hydraulic pump/motor supply auxiliary power, the hydraulic pump offers hydraulic oil for the pump/motor or charging hydraulic accumulator. The torque relationship is shown as [Eq. (4)]

$$T_d = T_{eng} i_g \eta_e \eta_t - T_P i_P \eta_P + T_{P/M} i_{P/M} \eta_{P/M} - T_{d,l}$$

Where $T_{P/M}$ is the torque of hydraulic pump/motor, $i_{P/M}$ is the velocity ratio of hydraulic pump/motor to the propeller shaft and $\eta_{P/M}$ is the efficiency of hydraulic pump/motor to the propeller shaft.

3.2.4. Accelerating/climbing condition

In order to enable the engine runs in a certain region or a smooth rise, the hydraulic pump does not work and the hydraulic pump/motor provides is shown as follows: the auxiliary power. The torque relationship. [Eq.(5)]

$$T_d = T_{eng} i_g \eta_e \eta_t + T_{P/M} i_{P/M} \eta_{P/M} - T_{d,l}$$

3.2.5. Regenerative braking condition

During braking, the hydraulic pump/motor works at pump condition, uses the braking energy to charge the hydraulic fluid from a low pressure hydraulic accumulator into a high-pressure accumulator, increasing the pressure of the nitrogen gas in the high-pressure accumulator. The regenerative braking force of hydraulic pump/motor is as follows: [Eq.(6)]

$$F_{P/M} = \frac{T_{P/M} i_0 i_{P/M} \eta_{P/M}}{r}$$

where i_0 is the final ratio of differential. It is designed and sized to capture braking energy from normal, moderate braking events and is supplemented by friction brakes and engine anti-trailer brakes for aggressive braking. The anti-trailer braking force of engine is determined by [Eq. (7)]

$$F_{eng} = \frac{I_e i_0^2 i_g^2}{r^2 \eta_r} \frac{dv}{dt} = \frac{I_e i_0^2 i_g^2}{r^2 \eta_r} z g$$

where I_e is the engine moment of inertia and z is the severity of braking.

4. OPTIMIZATION VIA DYNAMIC PROGRAMMING

Dynamic programming is a powerful tool to solve general dynamic optimization problems. In this study, the objective is to search for the optimal trajectories of control signals, including the engine command, hydraulic pump command and hydraulic pump/motor command to minimize the fuel consumption of the PHRBV over the whole driving cycle. i.e.[Eq.(8)]

$$J = \sum_{k=0}^{N-1} L(x(k), u(k)) + L_e(k) + L_b(k) + G(x(N))$$

where L is the fuel consumption over a time segment, N is the driving cycle length and x, u are the vectors of state variables and control signals respectively.

The objective function contains three components:

(1) The engine fuel consumption. This term only represents the fuel consumption assuming the engine is rotating in a steady state. [Eq.(9)]

$$L(x(k), u(k)) = P_e(k) * g_e(k) * t$$

(2) The second term is used to compensate the extra fuel consumption for the frequent start/stop engine and shifting gears. [Eq.(10) and (11)]

$$L_e(k) = \alpha (\text{sign}(n_e(k+1)) - \text{sign}(n_e(k)))$$

$$L_b(k) = \beta |i(k+1) - i(k)|$$

Frequent start/stop engine and shifting gears worsen fuel consumption and affect the ride comfortableness.

(3) In order to match the final value of accumulator SOC with its initial value, a penalty term is added. [Eq.(12)]

$$G = \gamma (\text{SOC}(N) - \text{SOC}(0))^2$$

Where α, β, γ are the weight coefficients.

Fig.4.1 presents the optimal trajectories of operating points of the engine and the vehicle operating modes over the NYCC duty cycle.

The accumulator pressure is characterized by large fluctuations due to high power flows through the system.

Large negative swings hydraulic pump/motor power indicate the effective capturing of braking energy. Since the Dynamic Programming algorithm is forward-looking, the resulting optimal control signals are not applicable in practice. By analyzing the DP results, the useful hints for deriving improved strategies can be practically implemented.

Clearly, at the beginning of each vehicle launch, the hydraulic pump/motor provides total propulsion power to avoid forcing the engine to work in the low speed/load region. With the exception of launch, the engine and pump/motor is used exclusively, because engine and hydraulic pump/motor have the characteristics of higher efficiencies at higher loads. Whenever the required power exceeds the maximum power provided by the pump/motor, the engine exclusive working mode is switched. In the meantime, the hydraulic pump is used to adjust engine working load through initiative charging pressure function and keep propulsion component at high load/high efficiency region. During the cruise speed stage, the pump/ motor is used to satisfy the total power demand whenever there is energy available in the accumulator, in the meantime, the frequent dynamical transitions between the various operation modes is needed to avoid.

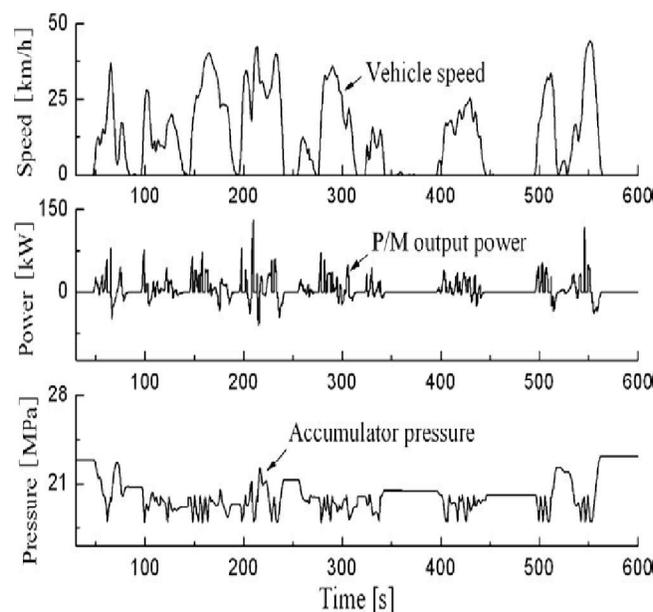


Fig. 4.1 Dynamic Programming results obtained over NYCC driving cycle.

5. FUZZY TORQUE CONTROL STRATEGY

Fuzzy torque control strategy uses fuzzy logic to build an energy distribution controller based on the useful hints derived from optimization results. The vehicle load of PHRBV varies in a wide range, if ignoring the change of vehicle load, the energy control strategy cannot realize

the fuel economy potential fullest. In the paper, the vehicle load is introduced as an input of the fuzzy logic controller through a sense proportion valve to determine the load condition of the vehicle. Appropriately reduce the pump charging torque and maximum charging pressure value when the load is small. Otherwise, enlarge the pump charging torque and maximum charging pressure value.

The three inputs of the fuzzy logic controller are the difference between the optimal torque (corresponding to the current engine rotation speed) and the vehicle required torque, the pressure (SOC) of the accumulator and the vehicle load. The difference between optimal torque and requirement torque (DT) is divided into five fuzzy subsets: {PL, PS, ZERO, NS, NL}. Similarly, the accumulator's SOC is divided into three fuzzy subsets: {HIGH, MED, LOW}, and the vehicle load is divided into three fuzzy subsets: {HIGH, MED, LOW}. The two outputs of the fuzzy logic controller are hydraulic pump/motor's output torque and the hydraulic ump's initiative charging pressure torque, while the fuzzy subsets are {LARGE, MED, SMALL, ZERO} and {LARGE, SMALL, ZERO}, respectively. Figs.5.1 and 5.2 present the input membership functions and the fuzzy logic block diagram.

Table 1 presents a list of if – then rules that represent the torque control strategy based on the analysis of DP optimization results.

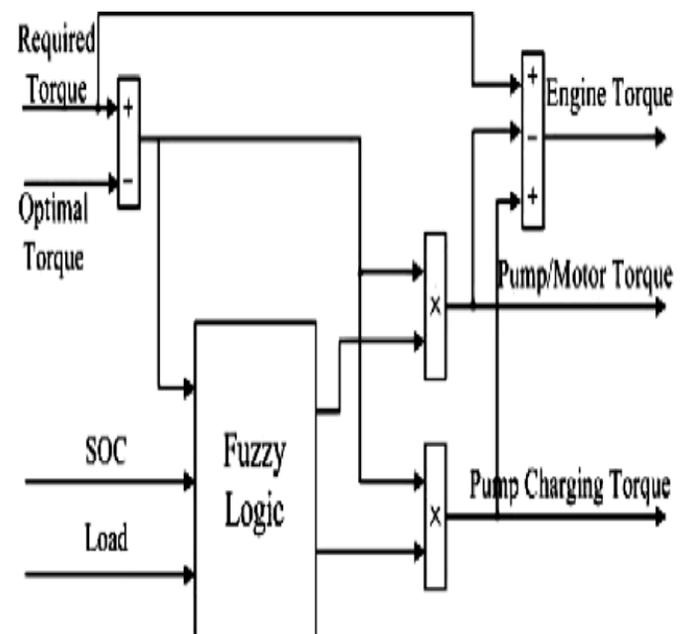


Fig. 5.1 Simplified block diagram of the fuzzy logic controller.

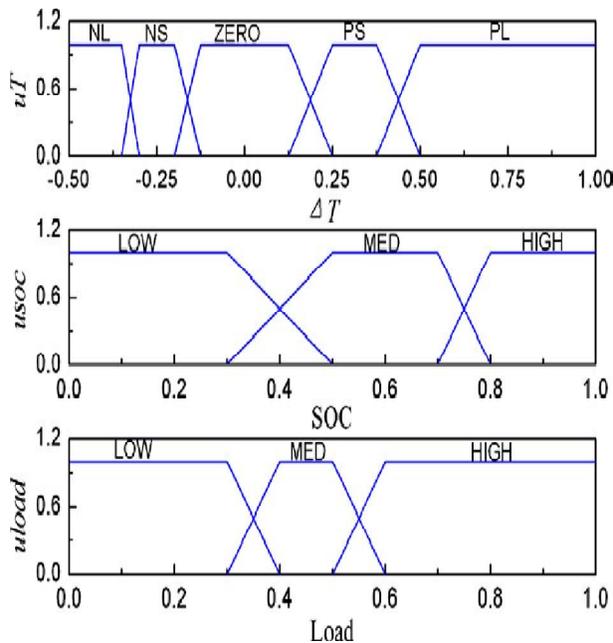


Fig. 5.2 Input membership functions for the scaled DT, SOC and Load.

TABLE 1
FUZZY CONTROL PRINCIPLE OF
TORQUE CONTROL STRATEGY

Conditions	Conclusions
If DT=NL & P = L = L, TP = Z	Then TP/M
If DT=NS & P = H = L, TP = Z	Then TP/M
If DT=NS & P =M = S, TP = Z	Then TP/M
If DT = Z = Z, TP = Z	Then TP/M
If DT = PL & P ffi L = L, TP = Z	Then TP/M
If DT = PS & P = H = L, TP = Z	Then TP/M
If DT = PS & P = M & load = L = Z, TP = S	Then TP/M
If DT = PS & P = M & load =M = Z, TP = S	Then TP/M
If DT = PS & P = M & load = H = Z, TP = L	Then TP/M
If DT = PS & P = L	Then TP/M

= Z, TP = L

The main principle of the fuzzy torque control rules are as follows: (1) DT is equal to zero, which shows that the engine is located in the highest efficiency region and working point need not to be adjusted; (2) DT is PL, which shows that the engine's efficiency is very low and it should be shut off to make the accumulator providing drive torque alone; (3) DT is PS, which shows that the engine is located in low efficiency region and initiative charging pressure function of hydraulic pump is used to adjust working point into the optimal region; (4) when DT is NL, it should meet the requirement of driving torque instead of economy; (5) when DT is NS, accumulator will help to reduce load.

6. SIMULATION RESEARCH

In order to check the validity of the fuzzy torque control strategy based on PHRBV, the PHRBV simulation model is implemented in SIMULINK. The simulation parameters of main components are listed in Table 2 and the vehicle configuration is shown in Fig. 1.

Traditional power-split strategy, optimal power management strategy and fuzzy torque control strategy are used to compare the improvement of fuel economy. The simulation results are shown in Figs. 6.1–6.3

TABLE 2
PARAMETERS OF THE VEHICLE AND
MAIN COMPONENT

Vehicle	
Wheel diameter	0.5 m
Rolling resistance	0.02
Aerodynamic coefficient	0.65
Frontal area	6.5 m ²
Total vehicle mass	14,310 kg
Accumulator system	
Volume	63 L
Max working pressure	32 MPa
Min working pressure	18MPa
Transmission	
Gear ratio	6.62, 3.99, 2.47, 1.55, 1
Main gear ratio	4.85

Hydraulic pump/motor

Type and displacement A4VG125
Max torque 795 nm

Hydraulic pump

Displacement 60

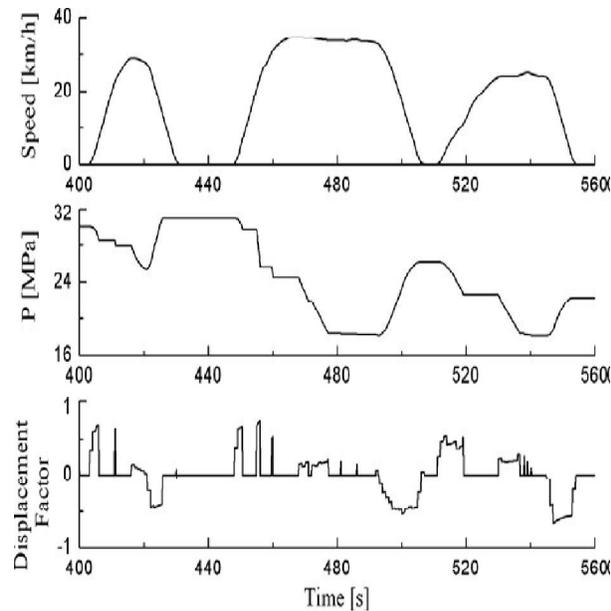


Fig. 6.1 Simulation results of power-split strategy

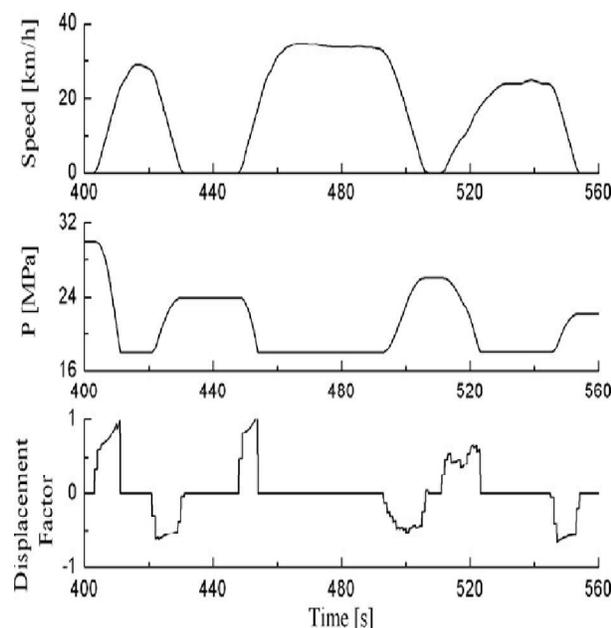


Fig. 6.2 Simulation results of optimal power management strategy.

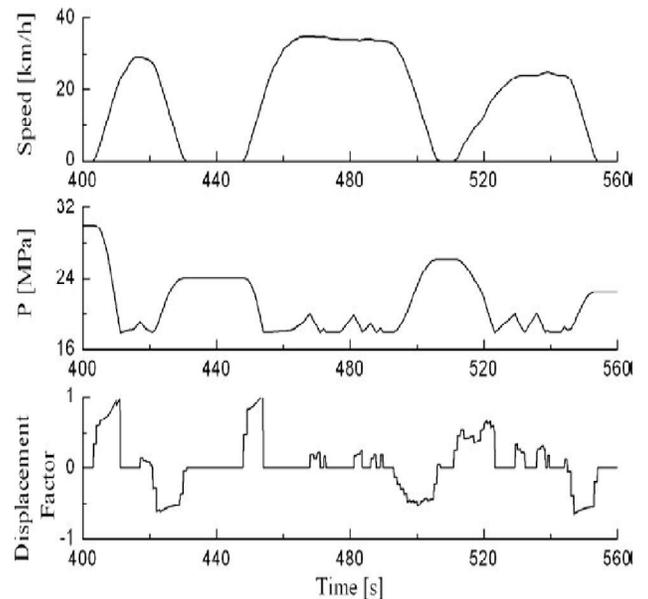


Fig. 6.3 Simulation results of fuzzy torque control strategy.

Figs. 6.1–6.3 show the differences between the power-split strategy, optimal power management strategy and fuzzy torque management strategy. Hybridization significantly improves vehicle's fuel economy over the city driving schedule. Hydraulic hybrid technology has unique characteristics of high power density compared to their electric counterparts, which enables regenerating and reusing significant amounts of braking energy. The large fluctuations of accumulator pressure demonstrated the effective regeneration and reuse of braking energy. Even using engine universal performance characteristics map to design control strategies (power-split strategy), the fuel economy improved 15.6% compared to the conventional vehicle. However, power-split strategy designed the energy distribution strategy only based on the static engine universal performance characteristics map and ignored the hydraulic hybrid propulsion system characteristics. Consequently, the energy distribution between primary and assistance sources is unreasonable. The regenerative braking ended earlier because the accumulator is fully charged. In addition, the frequent switching between engine mode and pump/motor mode worsens fuel consumption and ride comfortableness.

Optimal power management strategy proposed by used pump/motor to satisfy the total power demand, whenever there is energy available in the accumulator. If the power requirement is more than what pump/motor can provide, the engine will supplement the motor power. But the power management strategy ignored the relatively lower efficiency of engine when the propulsion

is switched to engine work mode at the high speed of the vehicle, so that the fuel economy potential can't be realized to its fullest.

Fuzzy logic controller is used to reasonably distribute the propulsion torque among the engine, hydraulic pump and hydraulic pump/motor. Hydraulic pump/motor is used to satisfy the total power demand, whenever there is energy available in the accumulator. The hydraulic pump is introduced to minimize the lower energy density disadvantages of hydraulic accumulator and move the engine working points into the highest efficiency region through initiative charging function. Hence, they enable a very dramatic increase of PHRBVs ability to realize its fuel saving potential. Implementation of fuzzy torque control strategy improved fuel economy to 32%.

Operating points of the engine controlled by power-split strategy were clustered in the mid/low load region (Fig.6.4). The fuzzy torque control strategy is able to move most points into the mid/high load zone characterized by highest efficiencies (Fig.6.5). In addition, frequent use of the motor for vehicle acceleration often depletes the energy in the accumulator, which prepares the system for the next regeneration event.

The impacts of load changing on the PHHV fuel economy are compared in Fig.6.6 Under 1015, ECE + EUDC, UDSS and HWFET driving cycles, the parallel hydraulic hybrid vehicle with considering the changes of load has higher fuel economy, which demonstrated the effectively of the fuzzy torque control strategy.

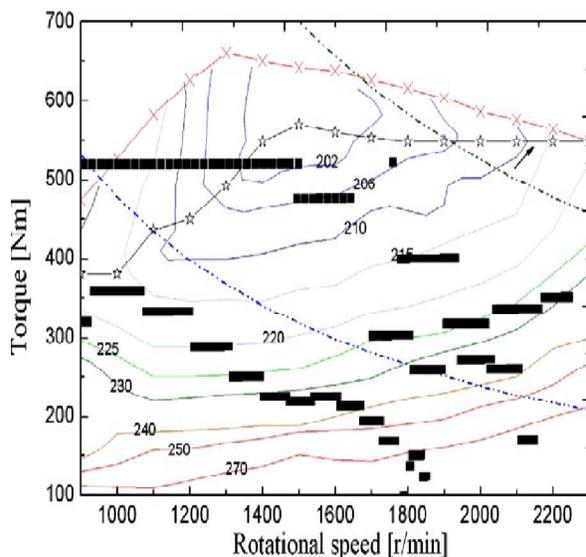


Fig. 6.4 Operating points of engine under power-split strategy.

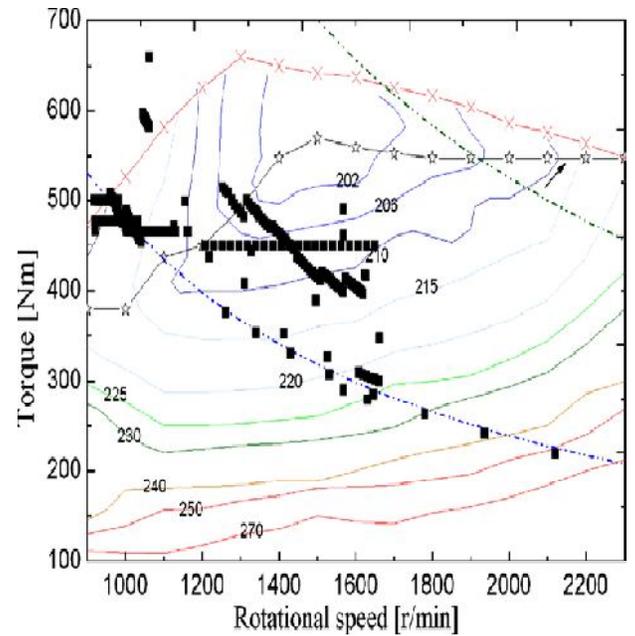


Fig.6.5 Operating points of engine under fuzzy torque control strategy.

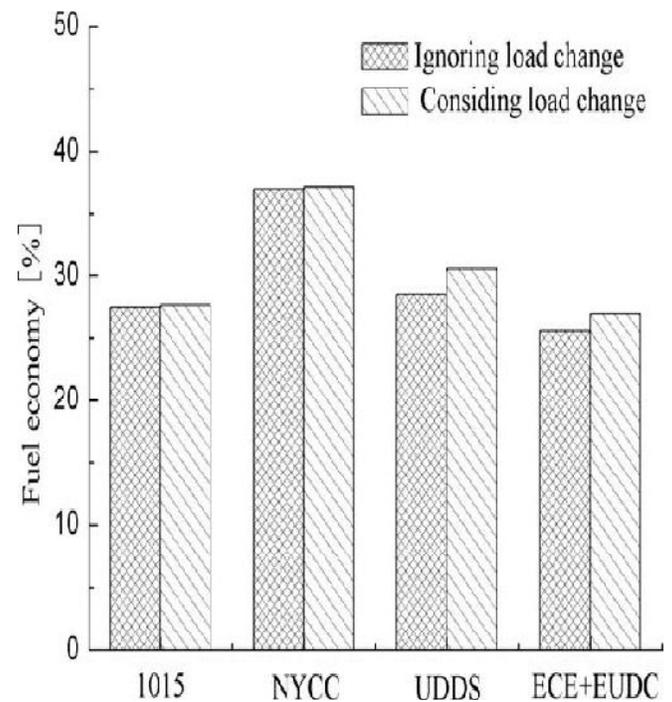


Fig. 6.6 Impacts of load changing on PHHV fuel economy

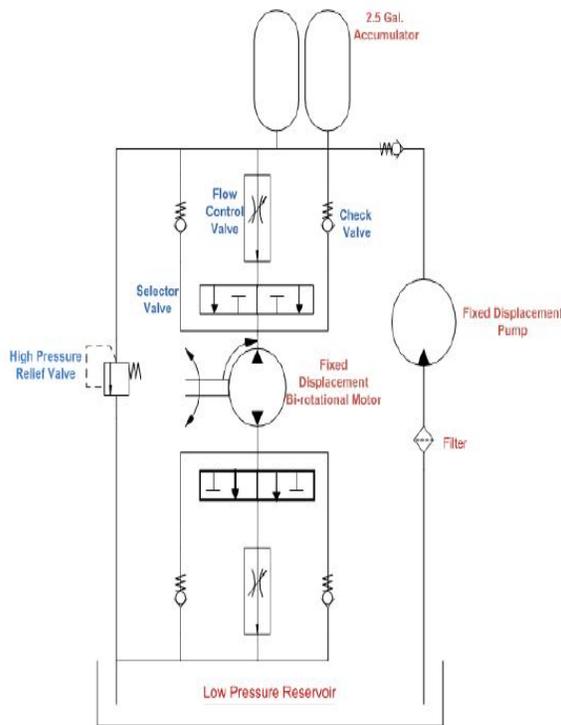


Fig. 6.7 Hydraulic system schematic

A simplified hydraulic system schematic is shown in Figure 6.7 for the three major driving scenarios: accelerating, braking, and coasting. The diagrams trace the fluids route through the system during each of the scenarios.

7. CONCLUSION

Hydraulic hybrid technology has the advantage of high power density and the ability to accept the high rates/high frequencies of charging and discharging, therefore it is well suited for off-road vehicles and heavy-duty trucks. But the lower energy density requires a special energy control strategy for PHRBV. In this study, a new type of configuration for PHRBV is presented. A fuzzy-based torque control strategy is built using the optimization results according to the torque distribution among the engine, hydraulic pump/motor and hydraulic pump, and the vehicle load changes is introduced to the fuzzy torque control strategy for realizing the fuel economy fullest. The simulation results show that the new configuration of PHRBV effectively improved the braking regenerative potential. The fuzzy torque control strategy reasonably distributed the propulsion energy between the power sources, improved the fuel economy and adaptability of different working conditions, and

minimized the disadvantage of accumulator's lower energy density, which provided a practical feasible method for improving fuel economy of the hydraulic hybrid vehicle.

REFERENCE

- [1] Wei YJ. Study on a new type of hydraulic hybrid sport utility vehicle. *China Mech Engg* 2006;17(15):1645-8(in Chinese)
- [2] Peng D, Yin CL, Zhang JW. Advanced braking control system for hybrid electric vehicle using fuzzy control logic. SAE paper No.2006-01-3583
- [3] Paul M, Jacek S. Development and simulation of a hydraulic hybrid powertrain for use in commercial heavy vehicle. SAE Paper No.2003-01-3370.
- [4] Hewko LO, Weber TR. Hydraulic energy storage based hybrid propulsion system for a terrestrial vehicle. *Energy Convers Engg Conf* 1990:99-105
- [5] Wu P, Luo N, Fronczak FJ, Beachle NH. Fuel economy and operating characteristics of a hydropneumatic energy storage auto-mobile. SAE paper 851678
- [6] Wu B, Lin CC, Filipi Z, Peng H, Assanis D. Optimal power management for a hydraulic hybrid delivery truck. *Vehicle Syst Dyn* 2004:23-40.
- [7] Lin CC, Peng P, Jessy W. power management strategy for a parallel hybrid electric truck. *IEEE Trans control Syst Technol* 2003(6):839-49.
- [8] Kim YJ, Filipi Z. Series hybrid propulsion for a light truck-optimization the thermostatic power management. SAE paper No.2007-24-0080.
- [9] Robyn AJ, Paul S, Steven B. Physical system model of a hybrid energy storage device for hybrid powertrain applications. SAE paper No. 2005-01-0810
- [10] Hiroki S, Shigeru I, Eitaro K. Study on hybrid vehicle using constant pressure hydraulic system with flywheel for energy storage. SAE paper No. 2004-01-3064.

A FPGA Implementation of Adaptive PPM Modulation Schemes for Wireless Optical Communication

Nirubama , Vijayakumar

Abstract— Wireless optical communication is the current research topic in high data rate data communication area. It is under research to utilize in under water sensor network communication and free space optical communication. Mainly PPM modulation scheme is used in FSO because of its high power efficiency, transmission efficiency and strong anti-jamming capability. One of the key difficulties of implementing PPM technique is that the receiver must be properly synchronized to align the local clock with the beginning of each symbol. Therefore, it is often implemented differentially as differential pulse-position modulation.

In this paper various PPM modulation schemes like single pulse position modulation, multi pulse position modulation, differential pulse position modulation schemes are going to be implemented in FPGA and its space requirement, time complexity and power consumption are going to be analyzed. This implementation is also concerned with an adaptive selection of modulation scheme based on channel condition, for this feedback received from receiver based on bit error rate.

Index Terms— adaptive modulation, wireless optical communication, fuzzy controller

Co-Author vijayakumar is currently working as assistant professor (sr.g) in electronic and communication engineering in SRM University, INDIA, E-mail: vijayakumar.p@ktr.srmuniv.ac.in

1 INTRODUCTION

Today's information economy depends on the transmission of data, voice and multimedia across telecommunication networks. Optical networks remain the most ideal medium for high-bandwidth communications for true connectivity. There are two distinct types of optical communications: Fiber optics (fiber-optic cable) and optical wireless systems based on free-space optics (FSO) technology. FSO is a line-of-sight technology that uses invisible beams of light to provide optical bandwidth connections that can send and receive voice, video, and data information. Free-space optics (FSO) refers to the transmission of modulated visible or infrared beams through the atmosphere to obtain broadband communications. Most frequently, laser beams are used, although non-lasing sources such as light-emitting diodes or IR-emitting diodes will serve the purpose.

The difference in fiber communication and FSO is that the energy beam is collimated and sent through clear air or space from the source to the destination, rather than guided through an optical fiber much used in the enterprise, mobile communication.

capable of sending up to 1.25 Gbps of data, voice, and video communications simultaneously through the air enabling fiber-optic connectivity without requiring physical fiber-optic cable.

Light travels through air faster than it does through glass, so it is fair to classify FSO technology as optical communications at the speed of light.

Compared to the radio frequency domain optical wireless communication offers much higher speeds and bit rates per watt

Modulation techniques have attracted increasing attention in optical wireless communications. Basic schemes such as on-off keying (OOK), pulse amplitude modulation (PAM) and pulse position modulation (PPM) have been validated as suitable for the optical wireless channel.

PPM is a well-known orthogonal modulation technique [2]. In L-PPM, a block of $\log_2 L$ input bits is mapped to one of L distinct waveforms, each including one "on" chip and L-1 "off" chips. A pulse $p(t)$ is transmitted during the "on" chip.

One of the principal advantages of PPM is that it is an M-ary modulation technique that can be implemented non-coherently, such that the receiver does not need to use a phase-locked loop (PLL) to track the phase of the carrier. This makes it a suitable candidate for optical communications systems, where coherent phase modulation and detection are difficult and extremely expensive. It improves transmission channel anti-jamming capability, consumes less transmission power and is average energy efficient.

The combined power and bandwidth requirements suggest that the basic modulation schemes cannot provide reliability when deployed in a real time channel, resulting in compromised system performance. So an adaptive selection of modulation scheme based on channel condition is the better solution for this. In this paper an adaptive PPM modulation scheme of FPGA implementation is presented.

Author nirubama is currently working as lecturer in electronic and communication engineering in Bharath University, India, E-mail: nirupsvijay@gmail.com

2 ADAPTIVE PPM MODULATION SCHEME

2.1 system outline

The selection of the modulation scheme is done according to power and bandwidth efficiency and the channel capacity. The power efficiency is a measure of how much signal power should be increased to achieve a particular BER (Bit Error Rate) for a given modulation scheme. Bandwidth efficiency is the ability to accommodate data within a limited bandwidth of a channel. It's also a tradeoff between data rate and pulse width. So in this project three different modulation schemes are implemented for wireless optical communication and based on bandwidth efficiency requirement or power efficient requirement on network side or application side we can configure any modulation scheme as current scheme

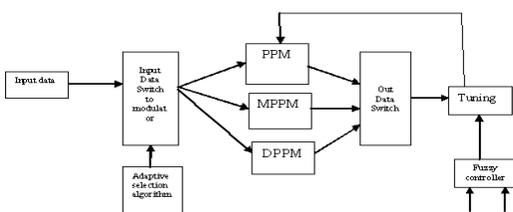


Fig.1 block diagram of adaptive PPM scheme

This scheme selects any one of the modulator dynamically based on mod set that is based on the scenario under which the system works. Various parameter like high band with requirement, constant data rate require and low power level in battery or in the device is used as key for this work.

Here single pulse PPM of the case 8-PPM is used Which will take 3 bit symbol as input and will produce 8-PPM output, Multi pulse PPM of dual pulse system is considered with slot size $M=5(\text{slots})$ which will take three bit input and a DPPM is generated by deleting all of the "off" chips following the "on" chip.

The realistic controller to change the state of the tuner block based on dynamic rapid Changing BER of the channel to maintain the constant target BER is implemented by fuzzy logic controller. Utilizing fuzzy systems in a dynamic control environment reduces potential stability problems. So this fuzzy based controller will stabilize the BER

2.2 method of adaptive selection

Adaptive selection scheme to select any one of the available implemented scheme is done here taking the pros and cons of the three PPM schemes. The single pulse PPM is low power efficient one but not bandwidth efficient .so it is not suitable for high bandwidth demand application and networks. It is configured as default type of modulation if there is no specific demand on application side, network side and receiver side.

High data rate demand application and for high data rate network scenario the DPPM is configured as modulation scheme but for constant bit rate receiver and network it is not a suitable scheme because of it's irregular data rate nature. So for such a

scenario either PPM or MPPM can be configured as current modulator. When current modulator configured is DPPM and suppose receiver report buffer overflow because of irregular data rate of DPPM, then MPPM is configured as current modulator to avoid the overflow. This selection scheme is given in the form of pseudo code in fig.2

```

If (mode="bandwidth efficient" or "application demand high data rate") then
    Modulation=DPPM;
Else if ( mode="low power mode" or mode="uniform data rate requirement or network="fixed through put network") then
    Modulation=PPM or MPPM;
    If (modulation=DPPM and receive buffer overflow="yes")
        then
            Modulation=MPPM;
        Else default:
            Modulation=PPM;
    
```

Fig.2: pseudocode for adaptive selection

2.3. Tuning for BER

Proposed modulation scheme takes the real time channel conditions into account, which is in form of current BER of the channel for tuning. By employing amplitude and position modulation selectively, a guaranteed system performance can be secured, without compromising power and bandwidth efficiency. This is also a new approach to realize reliable optical wireless links. A fuzzy logic control module has been developed to realize the tuning mechanism

2.4. Method of tuning

Fuzzy rule based tuning algorithm is used in this project. Utilizing fuzzy systems in a dynamic control environment reduces potential stability problems, and this is the benefit of applying fuzzy logic control over a modulation channel. Fuzzy logic (FL) is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth, that is truth values between "completely true" and "completely false". FL incorporates a simple, rule-based IF X AND Y THEN Z approach to a solving control problem rather than attempting to model a system mathematically

The standard definitions in fuzzy logic are:

$$\begin{aligned}
 \text{truth}(\text{not } x) &= 1.0 - \text{truth}(x) \\
 \text{truth}(x \text{ and } y) &= \text{minimum}(\text{truth}(x), \text{truth}(y)) \\
 \text{truth}(x \text{ or } y) &= \text{maximum}(\text{truth}(x), \text{truth}(y))
 \end{aligned}$$

Traditional control systems are based on mathematical models to define a relationship that transforms the desired state and ob-

served state of the system into inputs that will alter the future state of that system. Fuzzy machines work the same way, but the decision and the means of choosing that decision are replaced by fuzzy sets and rules. Fuzzy control, which directly uses fuzzy rules, has the purpose of influencing the operation of a system by changing inputs to that system via rules that model how the system operates

2.5. Fuzzy logic control algorithms

The System status can be modeled as membership functions in FL. The rules set in the FL system will come from the requirements of the ideal system. While BER varies across the range, FL can be used to set up rules according to different circumstances such as applications which need prolonged battery power, while the system power consumption is the main concern. In a system where data rate is more vital than others, such as in a bank or government offices, the selection of system status will be weighted more towards Rb to satisfy the need for high performance data transmission

A controller in the system needs to know the current BER level and needs to be able to set the state. Therefore, the controller's input will be the BER level difference (compared to 10⁻⁶, expressed in dB) and its output will be the rate, or the trend at which the M or n is changing. Since the BER level tends to oscillate around the desired level, it is also worth considering adding another input, the BER level's rate of change, to slow down the state change when the BER is close to the right level.

2.6 The rules for the fuzzy controller

1. If (BER is ok) then (state is no_change)
2. If (BER is low) then (slot/clock decreased_fast)
3. If (BER is high) then (slot/clock is increased_fast)
4. If (BER is ok) and (rate is negative), then (slot/clock is reduced_slow)
5. If (BER is ok) and (rate is positive), then (slot/clock is increased_slow)

According to the rules set above, the system performs a self-adaptation when BER degrades more than a certain threshold. Since high BER states are usually not acceptable from the communication system design point of view, the system states will be changed based on a calculation within available candidate states. For the tunable PPM scheme, M can be changed when the fuzzy controller has a positive output, and n changes when the output is negative. The operation could be inverted, dependent on the importance given to M and n.

3 FPGA IMPLEMENTATION

The FPGA is a form of highly configurable hardware. This is the reason why we implemented our modulation scheme in FPGA. More over, when sample rates grow above a few Mhz, a DSP processor or other processor has to work very hard to transfer the data without any loss. This is because the processor must use shared resources like memory busses, or even the processor core which can be prevented from taking interrupts for some time. An FPGA on the other hand with designed dedicates logic for receiving the data, can maintain high rates of I/O. since our modulation scheme of adaptive PPM is target for wireless optical communication which involves Giga bytes/sec data rate support the FPGA implementation will be more appropriate one. the other reason for going to FPGA implementation is FPGAs have product reliability and maintainability. My implementation uses altera Cyclone III EP3C16F484 FPGA.

4 RESULT AND ANALYSIS

The adaptive PPM modulation scheme is simulated for three input bit symbol using model sim simulator by setting the mode by means of three bit set as given in table I

TABLE I
MODE SETTING TO SELECT THREE DIFFERENT MODULATION SCHEMES

Mode state	modulation	Bit pattern
Low power mode/low bit rate mode	Single pulse PPM	000
High immune to noise or ISI mode/high power	Multi pulse PPM	001
High data rate/band with effient/non uniform data-rate mode	DPPM	010

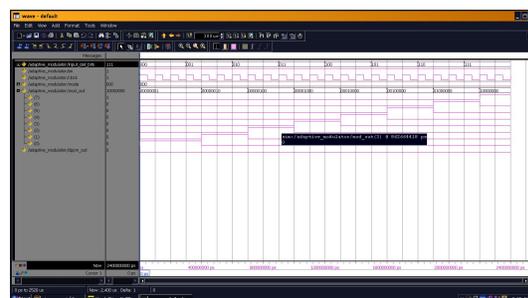


Fig.3 Screen shot out for low power mode modulator PPM

All input data combination is applied by setting mode-000 and the out put is generated. the modulated out signal mod_out is changing the pulse position in slots based on the input data change is shown in fig.1-3 here mod_out(0)-represent slot-0,mod_ou(1) represent slot-1 and so on.

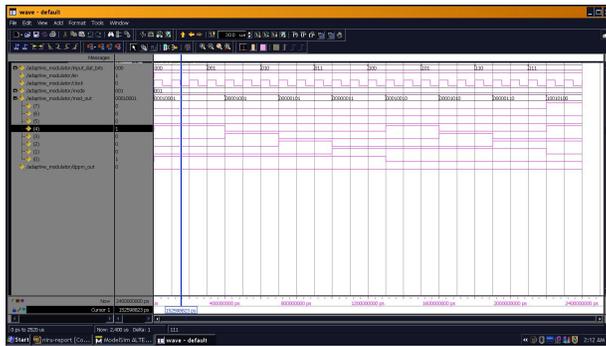


Fig.4 screen shot out for dual pulse PPM modulation

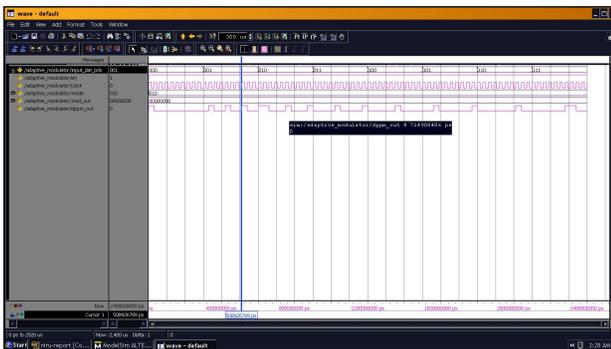


Fig:5 screen shot out for DPPM

4.1 RESORCE USAGE SUMMARY

- total logic element used-122
- total combinational functions-122
- total pin used -23
- Logic elements by mode
 - Normal mode 122
 - Arithmetic mode 0
- Total registers 0
- Dedicated logic registers 0
- I/O registers 0
- I/O pins 23
- Maximum fan-out node input_dat_bits[0]-input

- Maximum fan-out 22
- Total fan-out 466
- Average fan-out 2.77

The FPGA kit is programmed by generating the .SOF file by using quadras II software. Before generating .sof file the pin assignment is done by using pin planner tool of quadras II software

5 CONCLUSION

Reliability and multi network communication supports are two important requirements in wireless communication devices. By means of adaptive modulation scheme the multi network support feature can be implemented in communication systems. Reliability can be achieved by fine tuning the modulator parameter to achieve the target bit error rate .

Usually Only one type of modulator used in transmitter that will be suitable for the system that work in single environment with fixed channel condition and fixed data rate system. Recent days the communication device not only communicating in single network but in heterogeneous network environment. So the communication device have to re configure the modulation system that it uses based on the current network environment where it communicate .This multiple modulation system requirement of heterogeneous network environmental communication system leads to adaptive modulator design. This paper presents FPGA implementation of three different PPM modulation schemes of single pulse PPM, dual pulse PPM and DPPM for wireless optical communication that can be selectable by based on environment of communication network .DE0 Altera cyclone III EP3C16F484 FPGA kit used to demonstrate the scheme.

REFERENCES

- [1] [1]. Deng D., Chen S., Joos G., "FPGA Implementation of PWM Pattern Generator", IEEE Transactions on Communications, 2001; (40):225 – 230.
- [2] [2].J.M. D. Audeh, J. M. Kahn, and J. R. Barry, "Performance of pulse-position modulation on measured non-directed indoor infrared channels," Communications, IEEE Transactions on, vol. 44, pp. 654-659, 1996.
- [3] [3]. N. M. Aldibbiat, Z. Ghassemlooy, and R. McLaughlin, "Dual header pulse interval modulation for dispersive indoor optical wireless communication systems," IEEE Proceedings Circuits, Devices and Systems, vol. 149, no. 3, pp. 187-192, 2002
- [4] [4].J. J. Grubor, V. Langer, K.-D. , "Rate-adaptive multiple sub-carrier-based transmission for broadband infrared wireless communication," Optical Fiber Communication Conference, 2006 and the 2006 National Fiber Optic Engineers Conference, California, USA, pp. 10, 5-10 Mar
- [5] [5]. Nestor D. Chatzidihamantis,, Athanasios S. Lioumpas,, George K. Karagiannidis, and Shlomi Arnon, " Adaptive Subcarrier PSK Intensity Modulation in Free Space Optical Systems",under selection in IEEE transsation of communication system,feb 2010
- [6] [6].Aras S. Ameen, Ali Mohammad Noori & Lwa'a F. Abdul Ameer, Implementation of Programmable Pulse Timing Modulation Generator using FPGA, (KAJ)Kurdistan Academicians Journal, 2009/2010, 7, 8 (1) part A (99-106)

An Improvement of PSNR by Wavelet based True Compression of SAR Images

Senthilkumar.C, Dr.R.K.Gnanamurthy

Abstract - In Image Compression, the main intend is to shrink the number of bits required to represent an image by removing the spatial and spectral redundancies. The discrete wavelet transform has emerged as a popular technique for image compression. The result of the compression ratio changes as per the basis and type of wavelet used. In this paper focused on the selection of mother wavelet on the basis of nature of images, improve the quality as well as compression ratio extremely. We suggest the new technique, which is based on co-efficient thresholding. This method reduces the time complexity of wavelet packets decomposition. Our algorithm selects the sub-bands, which include significant information based on threshold entropy. The improved encoding technique of Huffman coding is suggested provides better results than existing compression methods.

Index Terms – Image, Compression, Co-efficient thresholding Method (CTM), Sub band coding, Wavelet, Huffman coding, True Compression.

1 INTRODUCTION

Image compression plays an essential role in several important and diverse applications like videoconferencing, remote sensing, medical imaging etc.,[1] These requirements are not fulfilled with old techniques of compression like Fourier Transform, Hadamard and Cosine Transform etc. The wavelet transform approach serves the purpose very efficiently. The basic idea behind the image compression is that in most of the images we find that their neighboring pixel. In this paper, we proposed compression scheme based on wavelet transform, which involves a discrete wavelet transform (DWT) for an image followed by quantization process for the wavelet coefficients after using a suitable bits allocation scheme [2]. The wavelet transform is one of the most exciting developments in the signal-processing field during the past decades. This is especially true when it is utilized in compressing images [3,4]. It is a well-established technique to compress independently the components of colored images and it has attracted a great interest in the area of image compression because of its excellent localization in both spatial and frequency domains.

The goal of true compression is to minimize the number of bits needed to represent it, while storing information of acceptable quality. Wavelets contribute to effective solutions for this problem. The complete chain of compression includes

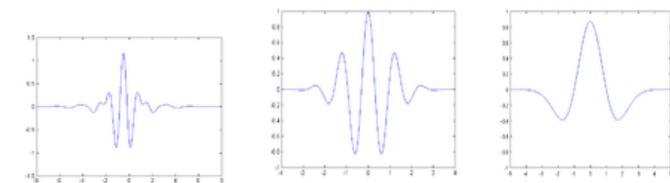
iterative phases of quantization, coding, and decoding, in addition to the wavelet processing.[5] A measure of achieved compression is given by the compression ratio (CR) and the Bit-Per-Pixel (BPP) ratio[6]. CR and BPP represent equivalent information. CR indicates that the compressed image is stored using CR of the initial storage size while BPP is the number. The goal of true compression is to minimize the number of bits needed to represent it, while storing information of acceptable quality. Wavelets contribute to effective solutions for this problem. The complete chain of compression includes iterative phases of quantization, coding, and decoding, in addition to the wavelet processing.[5] A measure of achieved compression is given by the compression ratio (CR) and the Bit-Per-Pixel (BPP) ratio[6]. CR and BPP represent equivalent information. CR indicates that the compressed image is stored using CR of the initial storage size while BPP is the number of bits used to store one pixel of the image. For a grayscale image the initial BPP is 8. For a true color image the initial BPP is 24, because 8 bits are used to encode each of the three colors (RGB color space) [7].The challenge of compression methods is to find the best compromise between a low compression ratio and a good perceptual result. The wavelet decomposes the image, and generates four different horizontal frequencies and vertical frequencies outputs. These outputs are referred as approximation, horizontal detail, vertical detail, and diagonal detail. The approximation contains low frequency horizontal and vertical components of the image. The decomposition procedure is repeated on the approximation sub-band to generate the next level of the decomposition, and so on. It is leading to well known pyramidal decomposition tree. Wavelets with many vanishing yield sparse decomposition [8] of piece wise smooth surface; therefore they provide a very appropriate tool to compactly code smooth images.

- Senthilkumar.C is currently pursuing research program in Digital image processing in Anna University f Tech.Coimbatore, India, PH-9842445859. E-mail: senthilkumar02@gmail.com
- Dr.R.K.Gnanamurthy is currently working as a Principal in Vivekanandha College of Engineering for Women, India, PH-9443734562. E-mail: princpalvcewengg@gmail.com

Further research has been done on still image compression and JPEG-2000 standard is established in 1992 and work on JPEG-2000 for coding of still images has been completed at end of year 2000. The JPEG 2000 standard employs wavelet for compression due to its merits in terms of scalability, localization and energy concentration [6, 7]. It also provides the user with many options to choose to achieve further compression. JPEG-2000 standard supports decomposition of all the sub bands at each level and hence requires full decomposition at a certain level. The compressed images look slightly washed-out, with less brilliant color. This problem appears to be worse in JPEG than in JPEG-2000 [9]. Both JPEG-2000 and JPEG operate in spectral domain, trying to represent the image as a sum of smooth oscillating waves. JPEG-2000 suffers from ringing and blurring artifacts. [9] The encoding phase of compression reduces the overall number of bits needed to represent the data set.

2. WAVELETS

A wavelet is a wave-like oscillation with an amplitude that starts out at zero, increases, and then decreases back to zero. It can typically be visualized as a "brief oscillation" like one might see recorded by a seismograph or heart monitor. Generally, wavelets are purposefully crafted to have specific properties that make them useful for signal processing. Wavelets can be combined, using a "shift, multiply and sum" technique called convolution [9], with portions of an unknown signal to extract information from the unknown signal. For example, a wavelet could be created to have a frequency of Middle C and a short duration of roughly a 32nd note. If this wavelet were to be convolved at periodic intervals with a signal created from the recording of a song, then the results of these convolutions would be useful for determining when the Middle C note was being played in the song. Mathematically, the wavelet will resonate if the unknown signal contains information of similar frequency - just as a tuning fork physically resonates with sound waves of its specific tuning frequency [11]. This concept



Meyer

Morlet

Mexican Hat

of resonance is at the core of many practical applications of wavelet theory.

As a mathematical tool, wavelets can be used to extract information from many different kinds of data, including - but certainly not limited to - audio signals and images. Sets of wavelets are generally needed to analyze data fully. A set of "com-

plementary" wavelets will deconstruct data without gaps or overlap so that the deconstruction process is mathematically reversible. Thus, sets of complementary wavelets are useful in wavelet based compression/decompression algorithms where it is desirable to recover the original information with minimal loss [10,11].

Wavelet theory is applicable to several subjects. All wavelet transforms may be considered forms of time-frequency representation for continuous-time (analog) signals and so are related to harmonic analysis. Almost all practically useful discrete wavelet transforms use discrete-time filter banks. These filter banks are called the wavelet and scaling coefficients in wavelets nomenclature. These filter banks may contain either finite impulse response (FIR) or infinite impulse response (IIR) filters. The wavelets forming a continuous wavelet transform (CWT) [12] are subject to the uncertainty principle of Fourier analysis respective sampling theory: Given a signal with some event in it, one cannot assign simultaneously an exact time and frequency response scale to that event. The product of the uncertainties of time and frequency response scale has a lower bound. Thus, in the scale gram of a continuous wavelet transform of this signal, such an event marks an entire region in the time-scale plane, instead of just one point. Also, discrete wavelet bases may be considered in the context of other forms of the uncertainty principle. In continuous wavelet transforms, a given signal of finite energy is projected on a continuous family of frequency bands (or similar subspaces of the L^p function space $L^2(\mathbb{R})$). For instance the signal may be represented on every frequency band of the form $[f,2f]$ for all positive frequencies $f>0$. Then, the original signal can be reconstructed by a suitable integration over all the resulting frequency components.

The frequency bands or subspaces (sub-bands) are scaled versions of a subspace at scale 1. This subspace in turn is in most situations generated by the shifts of one generating function $\psi \in L^2(\mathbb{R})$, the *mother wavelet*. For the example of the scale one frequency band [1,2] this function is

$$\psi(t) = 2 \operatorname{sinc}(2t) - \operatorname{sinc}(t) = \frac{\sin(2\pi t) - \sin(\pi t)}{\pi t}$$

with the (normalized) sinc function. Other example mother wavelets are:

The subspace of scale a or frequency band $[1/a, 2/a]$ is generated by the functions (sometimes called *child wavelets*)

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right)$$

where a is positive and defines the scale and b is any real number and defines the shift. The pair (a,b) defines a point in the right half plane $\mathbb{R}_+ \times \mathbb{R}$.

The projection of a function x onto the subspace of scale a then has the form

$$x_a(t) = \int_{\mathbb{R}} WT_{\psi}\{x\}(a, b) \cdot \psi_{a,b}(t) db$$

with wavelet coefficients

$$WT_{\psi}\{x\}(a, b) = \langle x, \psi_{a,b} \rangle = \int_{\mathbb{R}} x(t)\psi_{a,b}(t) dt$$

For the analysis of the signal x , one can assemble the wavelet coefficients into a scaleogram of the signal.

2.1. Discrete wavelet transforms

It is computationally impossible to analyze a signal using all wavelet coefficients, so one may wonder if it is sufficient to pick a discrete subset of the upper half plane to be able to reconstruct a signal from the corresponding wavelet coefficients. One such system is the affine system for some real parameters $a>1, b>0$. The corresponding discrete subset of the half plane consists of all the points $(a^m, n a^m b)$ with integers $m, n \in \mathbb{Z}$. The corresponding baby wavelets are now given as

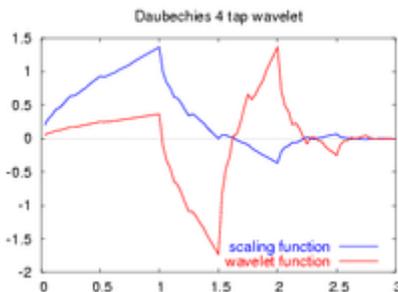
$$\psi_{m,n}(t) = a^{-m/2}\psi(a^{-m}t - nb).$$

A sufficient condition for the reconstruction of any signal x of finite energy by the formula

$$x(t) = \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} \langle x, \psi_{m,n} \rangle \cdot \psi_{m,n}(t)$$

is that the functions $\{\psi_{m,n} : m, n \in \mathbb{Z}\}$ form a tight frame of $L^2(\mathbb{R})$.

2.2. Multiresolution discrete wavelet transforms



D4 wavelet

In any discretized wavelet transform, there are only a

finite number of wavelet coefficients for each bounded rectangular region in the upper half plane. Still, each coefficient requires the evaluation of an integral. To avoid this numerical complexity, one needs one auxiliary function, the father wavelet $\phi \in L^2(\mathbb{R})$ [15]. Further, one has to restrict a to be an integer. A typical choice is $a=2$ and $b=1$. The most famous pair of father and mother wavelets is the daubechies 4 tap wavelet.

From the mother and father wavelets one constructs the subspaces

$$V_m = \text{span}(\phi_{m,n} : n \in \mathbb{Z}), \text{ where } \phi_{m,n}(t) = 2^{-m/2}\phi(2^{-m}t - n)$$

and

$$W_m = \text{span}(\psi_{m,n} : n \in \mathbb{Z}), \text{ where } \psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n).$$

From these one requires that the sequence

$$\{0\} \subset \dots \subset V_1 \subset V_0 \subset V_{-1} \subset \dots \subset L^2(\mathbb{R})$$

forms a multiresolution analysis of $L^2(\mathbb{R})$ and that the subspaces $\dots, W_1, W_0, W_{-1}, \dots$ are the orthogonal "differences" of the above sequence, that is, W_m is the orthogonal complement of V_m inside the subspace V_{m-1} . In analogy to the sampling theorem one may conclude that the space V_m with sampling distance 2^m more or less covers the frequency baseband from 0 to 2^{-m-1} . As orthogonal complement, W_m roughly covers the band $[2^{-m-1}, 2^{-m}]$.

From those inclusions and orthogonality relations follows the existence of sequences $h = \{h_n\}_{n \in \mathbb{Z}}$ and $g = \{g_n\}_{n \in \mathbb{Z}}$ that satisfy the identities

$$h_n = \langle \phi_{0,0}, \phi_{-1,n} \rangle \text{ and}$$

$$\phi(t) = \sqrt{2} \sum_{n \in \mathbb{Z}} h_n \phi(2t - n)$$

and

$$g_n = \langle \psi_{0,0}, \phi_{-1,n} \rangle \text{ and}$$

$$\psi(t) = \sqrt{2} \sum_{n \in \mathbb{Z}} g_n \phi(2t - n)$$

3. Wavelet and Wavelet Packet

In order to represent complex signals efficiently, a basis func-

tion should be localized in both time and frequency domains. The wavelet function is localized in time domain as well as in frequency domain, and it is a function of variable parameters. The wavelet decomposes the image, and generates four different horizontal frequencies and vertical frequencies outputs. These outputs are referred as approximation, horizontal detail, vertical detail, and diagonal detail. The approximation contains low frequency horizontal and vertical components of the image. The decomposition procedure is repeated on the approximation sub-band to generate the next level of the decomposition, and so on. It is leading to well known pyramidal decomposition tree. Wavelets with many vanishing yield sparse decomposition of piecewise smooth surface; therefore they provide a very appropriate tool to compactly code smooth images. Wavelets however, are ill suited to represent oscillatory patterns [13, 14]. A special from a texture, oscillating variations, rapid variations in the intensity can only be described by the small-scale wavelet coefficients. Unfortunately, these small-scale coefficients carry very little energy, and are often quantized to zero even at high bit rate [16] up to third level.

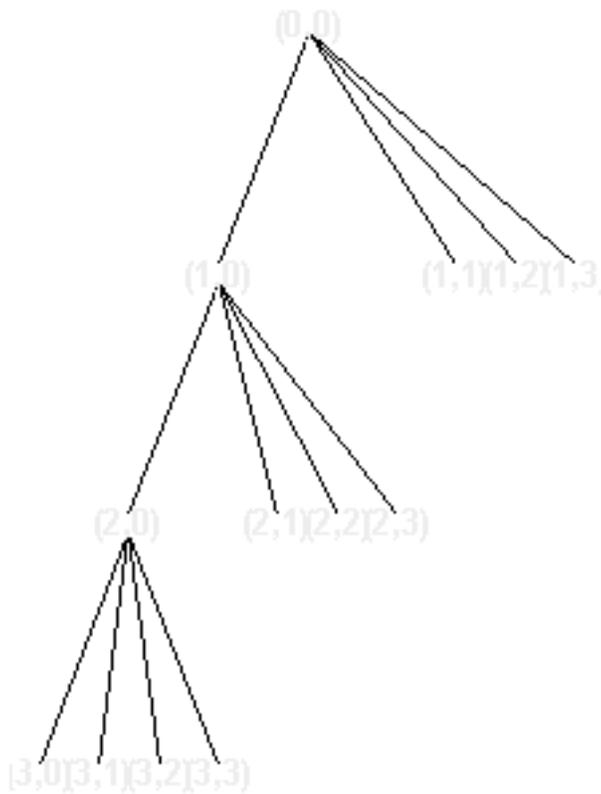


Figure 1. The-tree structure of wavelet decomposition

LL1LL2	LL1HL2	HL1LL2	HL1HL2
LL1LH2	LL1HH2	HL1LH2	HL1HH2
LH1LL2	LH1LH2	HH1LL2	HH1HL2
LH1LH2	LH1HH2	HH1LH2	HH1HH2

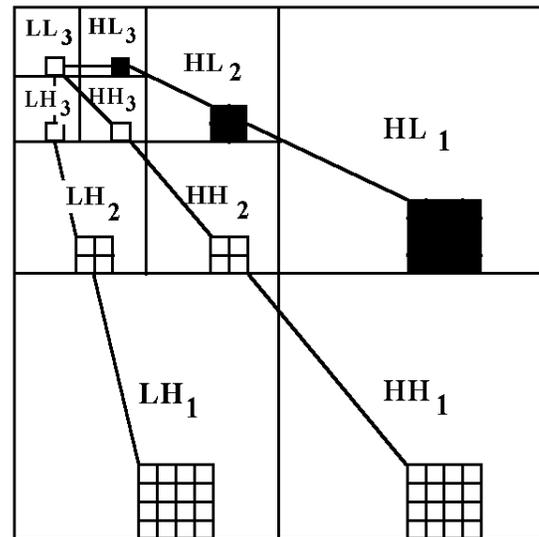
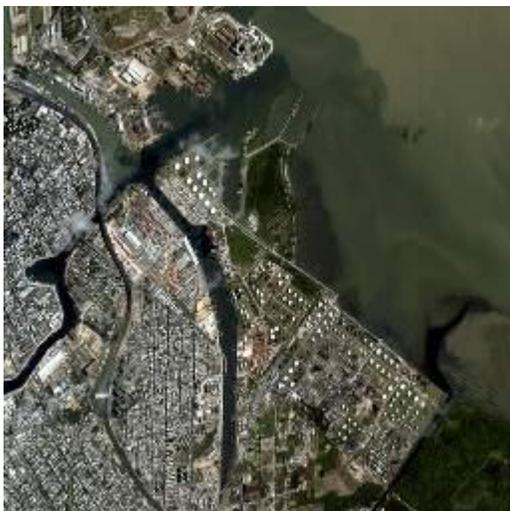


Figure 2. The structure of two level decomposition of Wavelet packet

4. Results and Conclusions

The proposed method is applied separately on the luminance component and both chrominance components of satellite colored image which has 24b/p and its size is 256X256 pixel(see figure 1). All involved parameters, implied within this scheme, were utilized as control parameters to investigate the compression performance of the proposed method [13]. These parameters are; Index of weighting type, inclusion factor, and number of layers. The effects of these parameters were investigated by considering several cases within the allowable range of their values, as shown in tables (1 -2) , while figure (4) illustrates the effect of the control parameters on the compression performance (including the quality of produced luminance component and the produces compression ratio). Figures, (3) , presents the reconstructed RGB image for YIQ color model. From the above results some remarks related to the behavior and performance of the proposed compression method could be presented as follows.



Number of layers =3, Index of weighting type =1
Inclusion factor =1, C.R.=11.515 ,PSNR = 19.627

Fig. (4)The original colored image

Table 1

Test color Image	Index of weighting type	If	C.R.	PSNR(dB)
Satellite	0	1	4.137	23.658
	1	1	10.684	19.670
	2	1	10.684	19.670
	0	1.6	14.595	17.272
	1	1.6	11.757	19.157
	2	1.6	11.757	19.157
	0	2	15.194	17.182
	1	2	12.206	19.078
	2	2	12.206	19.078



Number of layers =2, Index of weighting type =1
Inclusion Factor =1, C.R.=10.684, PSNR = 19.670

REFERENCES

- [1] Menegaz, G., L. Grewe and J.P. Thiran, 2000 Multirate Coding of 3D Medical Data. In proceedings of the 2000 International Conference on Image Processing, IEEE, 3: 656-659.
- [2] Misiti, M., Y. Misiti, G. Oppenheim, J.-M. Poggi (2007), "Wavelets and their applications", ISTE DSP Series.
- [3] Paul S. Addison, *The Illustrated Wavelet Transform Handbook*, Institute of Physics, 2002, ISBN 0-7503-0692-0.
- [4] Ali Akansu and Richard Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands, Wavelets*, Academic Press, 1992, ISBN 0-12-047140-X.
- [5] B. Boashash, editor, "Time-Frequency Signal Analysis and Processing – A Comprehensive Reference", Elsevier Science, Oxford, 2003, ISBN 0-08-044335-4.
- [6] Tony F. Chan and Jackie (Jianhong) Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, Society of Applied Mathematics, ISBN 089871589X (2005).

- [7] Ingrid Daubechies, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, 1992, ISBN 0-89871-274-2.
- [8] Ramazan Gençay, Faruk Selçuk and Brandon Whitcher, *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, Academic Press, 2001, ISBN 0-12-279670-5.
- [9] Haar A., *Zur Theorie der orthogonalen Funktionensysteme*, Mathematische Annalen, **69**, pp 331–371, 1910.
- [10] Barbara Burke Hubbard, "The World According to Wavelets: The Story of a Mathematical Technique in the Making", AK Peters Ltd, 1998, ISBN 1568810725, ISBN 978-1568810720.
- [11] Gerald Kaiser, *A Friendly Guide to Wavelets*, Birkhauser, 1994, ISBN 0-8176-3711-7.
- [12] Stéphane Mallat, "A wavelet tour of signal processing" 2nd Edition, Academic Press, 1999, ISBN 0-12-466606-x
- [13] Donald B. Percival and Andrew T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, 2000, ISBN 0-5216-8508-7
- [14] Press, WH; Teukolsky, SA; Vetterling, WT; Flannery, BP (2007), "Section 13.10. Wavelet Transforms", *Numerical Recipes: The Art of Scientific Computing* (3rd ed.), New York: Cambridge University Press, ISBN 978-0-521-88068-8, <http://apps.nrbook.com/empanel/index.html#pg=699>
- [15] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, 1993, ISBN 0-13-605718-7
- [16] Mladen Victor Wickerhauser, *Adapted Wavelet Analysis From Theory to Software*, A K Peters Ltd, 1994, ISBN 1-56881-041-5.

Low Power Soc Communication Using Steiner Graph AMBA Architecture

Krishna Kumar.S

Abstract—Reduction in size of a chip leads to more sophisticated design, but the system components must be designed to operate with considerably low power. Altering the data and control path in a design will reduce the power dissipation in a chip to a greater extent. The limitation of a normal bus design is that data or control values will be transmitted to all the Soc components invariably, thus the chip select control lies within the components, either single or many circuit will be activated with the reference to control signal but the transmission cost is high due to the unnecessary passage of signal in the bus path. Steiner graph method of predicting the best bus routing path combined with Gated clock tree structure, which will further make the design more flexible. The bus implementation is done by using AHB Bus protocol for its robust and flexible design in nature. Thus a low power operating Soc can be designed and implemented without modifying the structure of component but by altering bus path structure. The level of power consumption can be further reduced by modifying the structure of the components too.

Index Terms— Communication Architecture, Low Power Design, System-On-Chip, Steiner Graph, Clock Gating.

1 INTRODUCTION

The feature size of process technology scales down. The SOC are capable of integrating more and gaining higher complexity. The clock frequency is limiting due to power and thermal limitations, better performance is achieved through parallelism. The on-chip communication become critical in future due to communication latency and bandwidth components become a task the power consumption on inter component communication has scaled up to a significant level. The existing on-chip bus standards can provide an interface for IP developers and for system designers. Compared to network-on-chip type of communication buses are small on silicon footprint, fast in terms of latency and easy to implement[2]. Further the implementation enabling the designers to apply various optimizations for best performance with available resources. The present bus architecture are not power efficient to transfer on bus lines. When high bandwidth is required on these buses the wire efficiency will be low which limits system bandwidth capacity.

The proposed synthesis scheme for on-chip buses eliminates the disadvantages in existing bus system without changing the existing protocols and component interfaces are based on shortest path Steiner graphs bus lines are maximized which needs no modified design of system components and IP modules. Low power is maintained with reduced routing resources. The technology trend welcomes this physical synthesis scheme for bringing a large improvement on power performance based on current on-chip buses and bus matrices. To evaluate the system punctuality a communication constraint graph is extracted from a specific application on which topology configuration has an estimated performance by analysis and simulation [3]. An elaborate power analysis on AMBA on-chip bus is performed. where the detailed decomposition of power consumed by system components is obtained by simulation on NEC'S gate-level power estimator. Clock gating is nowadays used to reduce dynamic power and power gating is used to avoid unnecessary static power. In bus communications, a large part of power is consumed on the wires of bus lines.

•krishna kumar.S is currently pursuing masters degree program in Embedded system technologies in Veltech multitech Dr.RR and Dr.SR Engineering college affiliated to Annauniversity, India, PH-+919943812995. E-mail:Krishna.veltech@gmail.com

IJSER © 2012

<http://www.ijser.org>

A power performance tradeoff is analyzed on bus matrices, where a bus matrix is composed of a set of tree structured buses[4]. These structures from trees to graphs, using Steiner graph connection for optimization of bus gating to minimize the communication power. On the power side communication power depends on the wire capacitance in data transaction and the performance side, delay is dominated by signal propagation distance bandwidth is limited by routing congestion and resulted in power thermal constraints.

The rest of the paper is organized as follows. Section2 first introduces Bus matrix design. Next, the proposed power reduction using bus gating technique combined with Steiner graphs is described in Section3. Section 4 then presents Implementation. Section 5 then concludes this paper.

2 BUS MATRIX DESIGN

In this paper we look at bus matrix based communication architectures which are currently being considered by designers to meet the high bandwidth requirements of modern Soc systems. Fig.1 shows an example of a three-master seven-slave AMBA bus matrix architecture for a dual ARM processor based networking subsystem application.

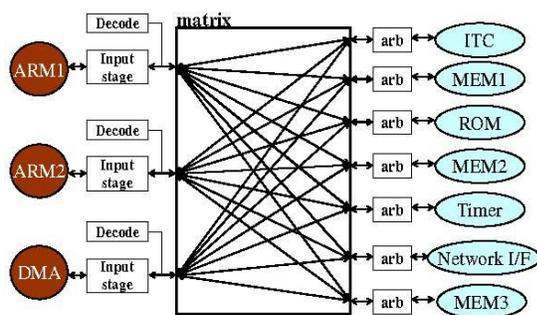


Fig.1. Full bus matrix architecture

A bus matrix consists of several busses in parallel which can support concurrent high bandwidth data streams. The Input stage is used to handle interrupted bursts, and to register and hold incoming transfers if receiving slaves cannot accept them immediately[6]. The Decode stage generates select signal for appropriate slaves. Unlike in traditional shared bus architectures, arbitration in a bus matrix is not centralized, but rather

distributed so that every slave has its own arbitration. One drawback of the full bus matrix structure shown in Fig. 1 is that it connects every master to every slave in the system, resulting in a prohibitively large number of busses in the matrix. The excessive wire congestion can make it practically impossible to route and achieve timing closure for the design.

3 POWER REDUCTION USING BUS GATING TECHNIQUE

Standard on-chip buses like AMBA were designed to enable fast and convenient integration of system components into the Soc, where simplicity is one of the major objectives. When the bus power consumption comes to a significant level that we cannot afford to ignore, power optimization will be desirable. We introduce a “bus gating” technique to minimize the power on bus lines with a small compromise on design simplicity[5]. The power efficiency of bus architecture like is low because the bus lines from masters to slaves are connecting all the slave devices by a single large wire net. The same is on slave-to-master connections. While the communication is one-to-one, the signals are sent to all the receivers regardless of whether they are needed, which results in wasted dynamic power on bus wires and component interfaces as shown in fig2. For tree structured buses distributing the multiplexer and de-multiplexer into the wire net helps to save both power and wires.

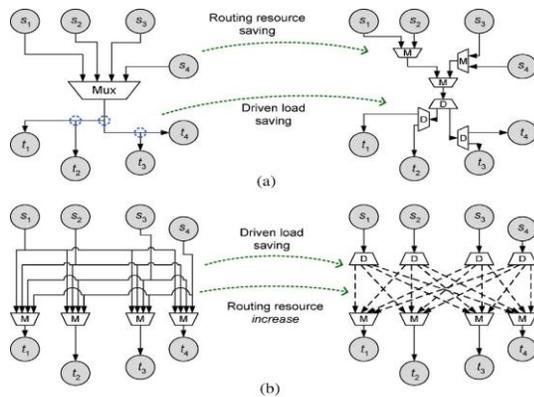


Fig.2. Bus gating using distributed mux and de-mux.(a)On single bus.(b)On bus matrix.

Arborescence is a directed tree such that every root-to-leaf path is shortest. On the receiver side with distributed de-multiplexers, the bus lines change from a rectilinear Steiner minimum tree to minimum rectilinear Steiner arborescence (MRSA).

By the research this change increases the wire length by only 2–4% on average. So the total bus wire length can be reduced by the distributing the multiplexer/de-multiplexers, while the dynamic power can also be reduced at the same time[1]. There is a small control overhead for sending the signals over the arborescence, but compared to the bus width and data throughput, this dynamic power overhead is negligible.

Based on the same tree topology, effective bus gating can be applied by distributing the control over the entire tree. On bus matrices however simply adding de-multiplexers may increase the total wire length, because when the number of master-to-slave paths becomes large, each path will need its own bus wires.

To reduce wire length in the bus matrix, also to further reduce power on the basic bus, we adopt the structures of Steiner graphs. A Steiner graph is a generalization of Steiner trees, without the limitation of tree structure that there is only one root placed at a certain point, which cannot be on the shortest path of every connection[7].

By removing the constraint of tree topologies, we gain higher freedom to choose shortest paths for reduced power on data transactions, and to let the paths share wires for reduced routing congestion. As defined in for an un weighted graph $G=(V,E)$, $G=(V,E, \omega)$ is a Steiner graph of G if $V \subseteq V$ and for any pair of vertices $u, w \in V$, the distance between them in G is at least the distance between them in G . Fig.3 shows a Steiner graph of G with $V=\{s1,s2,t1,t2\}$, $E=\{(s1,t1),(s1,t2),(s2,t1),(s2,t2)\}$ with each edge weighted 1, and its implementation as a bus or bus matrix.

This graph is minimal in terms of total wire length. Moreover, every edge in E has a path in G with minimum length, i.e., the path length equals the Manhattan distance between the two vertices.

In this way, each data transaction involves minimal wires, leading to minimal dynamic power on bus lines.

Shortest-path steiner graphs have advantages on power efficiency as shown above. . Naturally graph structures also have advantage on communication bandwidth over trees[7].

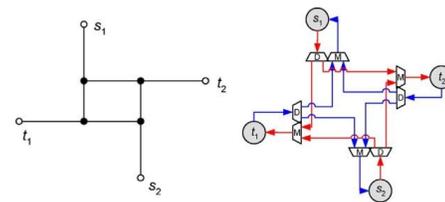


Fig. 3. Shortest-path Steiner graph G'' and its Bus implementation

Our objective is bus gating and bus matrix synthesis is to perform a balanced optimization on power and bandwidth even when available routing resource is limited.

4 IMPLEMENTATION

A) Steiner Graph

The Steiner tree problem or the minimum Steiner tree problem is a problem in combinational optimization which may be formulated in a

number of settings with the common part being that it is required to find the shortest interconnect for a given set of objects. The Steiner tree problem is superficially similar to the minimum spanning tree problem given a set V of points (vertices), interconnect them by a network graph of shortest length, where the length is the sum of the lengths of all edges. The difference between the Steiner tree problem and the minimum spanning tree problem is that, in the Steiner tree problem, extra intermediate vertices and edges may be added to the graph in order to reduce the length of the spanning tree[8]. These new vertices introduced to decrease the total length of connection are known as Steiner points or Steiner vertices.

It has been proved that the resulting connection is a tree known as the Steiner tree. There may be several Steiner trees for a given set of initial vertices. The Steiner tree problem has applications in circuit layout or network design.

B) Clocked Gate

A clock gated is an justify architecture in which the unnecessary path or power losses will be controlled with much effectively like utilizing gated driver tree structure which can exactly control the transmission path inside a bus. So that it will not transmit unnecessary nodes. It named as clock gated driver because the clock also will not be transmitted to other nodes[5]. The clock utilization for transmission be through the destination not unnecessary node path, in further we will reduce the power consumption in existing work. To save area, the memory module of a delay buffer is often in the form of static ram array with input/output data bus. Special read/write circuitry, such as a sense amplifier, is needed for fast and low-power operations. However of all the memory cells, only two words will be activated. Driving the input signal all the way to all the memory cells seems to be waste of power. This can be avoided in Steiner graph combined with clocked gating technique.

C) Steiner Graph Combined With Clock Tree Gating

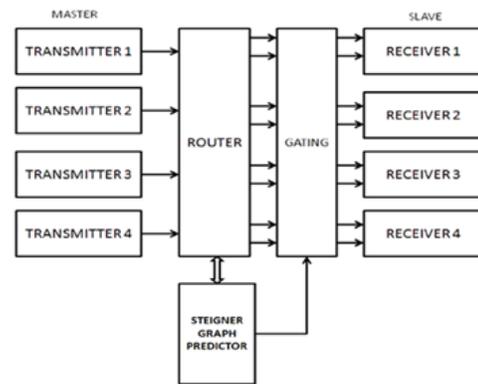


Fig.4. Steiner graph combined with clock tree gating

The bus architecture has a limitation of transmitting or passing the signal throughout network structure so all the nodes in the network will receive the input data and the particular node which is used or which decide as the destination will assign the data and will acknowledge the data other nodes will simply ignore the data which does not have the destination address so there will be unnecessary path delay.

Unnecessary path loss in transmitting the data through the unnecessary nodes [5]-[9]. So in our proposed design in fig.4 we are modifying the bus structure so that it can deliver only to the exact destination node. It will reduce the power loss, path metric and have lot of advantage through bus design. The prediction of the exact node or shortest path of node will be using Steiner graph which is used to find the exact path metric and used to find the exact bus destination occurs. Some of the advantages of proposed system in clock tree architecture will be introduced to easily access of a particular memory block. We can easily navigate the Bus structure of Bus path to select and deselect one or more path parallel and Bus utilization is reduced and so power is minimized.

5 CONCLUSION

In this paper we propose a physical synthesis scheme for on-chip buses and bus matrices to

minimize the power consumption, without changing the interface or arbitration protocols. By using a bus gating technique, data transactions can take shortest paths on chip, reducing the power consumption of bus wires to minimal. Routing resource and bandwidth capacity are also optimized by the construction of a shortest-path Steiner graph, wire sharing among multiple data transactions and wire reduction heuristics on the Steiner graph. Experiments indicate that the gated bus from our synthesis flow can save more than 90% dynamic power on average data transactions in current AMBA bus systems.

REFERENCES

- [1] Renshen Wang, Yulei Zhang, Nan-Chi Chou, Evageline, Ronald "Bus matrix synthesis based on steiner graph for power efficient soc communication," in IEEE Transaction. Computer-Aided Design, 2011, pp. 167-179.
- [2] W. Dally and B. Towles, "Route packets, not wires: On-chip interconnection network," in Proc. ACM/IEEE Des. Autom. Conf., Jun. 2001, pp. 684-689.
- [3] K.Lahiri, A.Raghunathan, and S. Dey, "Efficient exploration of the SoC communication architecture design space," in Proc. Int. Conf. Comput.-Aided Design, 2000, pp. 424-430.
- [4] S.Pasricha, Y.-H. Park, F. J. Kurdahi, and N. Dutt, "System-level power performance tradeoffs in bus matrix communication architecture synthesis," in Proc. Int. Conf. Hardw.-Softw. Codesign Syst. Synthesis, 2006, pp. 300-305.
- [5] L.A.Ca, Q.Wu, M.Pedram and X.Wu, "Clock-gating and its application to low power design of sequential circuits," in Proc. IEEE Custom Integr. Circuits Conf., vol. 47. Mar. 2000, pp. 415-420.
- [6] Sudeep Pasricha, Nikil Dutt and Mohamed Ben-Romdhane, "Bus matrix communication architecture synthesis," in CECS Conf, 2005
- [7] W.Shi and S.Chen, "The rectilinear Steiner arborescence problem is np complete," in Proc. ACM-SIAM Symp. Discrete Algorithms, 2000, pp. 780-787.
- [8] R.Wang ,N.C.Chou, B.Salefski, and C.K. Cheng, "Low power gated bus synthesis using shortest-path Steiner graph for system-on-chip communications," in Proc. ACM/IEEE Des. Autom. Conf., Jul. 2009, pp. 166-171.
- [9] Po-Chung, Jing-Siang Jhuang, Pei-Yun Tsai, "A low power delay buffer using gated driver tree", in IEEE Transaction. Very Large Scale Integration, 2009

Geo-processing using Oracle Spatial Geo Database

¹Manoj Pandya, ¹Pooja Nair, ¹Parthi Gandhi, ¹Shubhada Pareek, ²Dr. Bijendra Agarwal

Abstract - In the last two decades, Geographic Information Systems (GIS) has been widely developed and is applied to various fields, which include resources management, environment management, prevention of disaster, planning area, education and national defense etc in many sections and subjects. However traditional GIS application systems can't interact with each other and was considered as an isolated island of information facing problem of interoperability. This is because of different data models adopted by different GIS applications. The spatial data interoperability concept brought forward a huge amount of geospatial information for efficacious management, sharing and increase in value usage. OGC standards are developed in a unique consensus process supported by the OGC's industry, government and academic members to enable geo-processing technologies to interoperate, or "plug and play".

Index Terms – Geo Database, Geo processing, Oracle Spatial, OGC, 3D Rendering, Extrude, GIS

1. INTRODUCTION

Within an enterprise, various software packages are used and in many cases they can't talk with each other. This situation is ubiquitous and also exists between different departments. An Open GIS Consortium (OGC) is an international industry consortium of more than 400 companies, government agencies and universities participating in a consensus process to develop publicly available interface standards. Within a local government or enterprise, the spatial data is centrally stored. Interoperable metadata enables organizations to use the right tool for the job while eliminating complicated data transfers and multiple copies of the same data throughout the enterprise or department. The application server is a mediation system, this model uses oracle application server as the mediation system, and through the application server the application client sends WMS or WFS request and get the map server for background application. The three-tier structure model exposes a GIS portal which is an online GIS for outside of the department. Any client can request the server if it accords with WMS or WFS specification.

2 DATABASE PLATFORMS

There are several database platforms which provide the spatial support. Oracle® is matured to support OGC

standards. However there are other database platforms available like Microsoft® SQL Server 2008, IBM DB2 etc. Open Source database platforms are also available in the realm of proprietary products. MySQL and PostGRES with PostGIS extension can server large scale Geo-Database.

3 ORACLE SPATIAL DATA MODEL

Oracle® offers a spatial data model that provides basic enterprise access to geospatial information. This spatial data model provides a standard structure for point, line, and area features. Oracle Spatial 10g XE (Express Edition) is used in this prototype model. This database platform is free to use upto 4 GB of data. However enterprise version has no limit. Basic DDL (CREATE, ALTER, DROP) and DML (INSERT, UPDATE, DELETE) statements can be used for Database Management. With Spatial, the geometric description of a spatial object is stored in a single row, in a single column of object type SDO_GEOMETRY in a user-defined table.

- Manoj Pandya is currently working as a senior project scientist in Bhaskaracharya Institute for Space Applications and Geo-Informatics (BISAG), Gandhinagar, Gujarat, India PIN 382007 E-mail: mjpandya@gmail.com
- Pooja Nair, Parthi Gandhi, Shubhada Pareek are pursuing project at BISAG
- Dr. Bijendra Agarwal is Head Of Department (HOD), VJMS College, VADU, Kadi, Mahesana, Gujarat, India

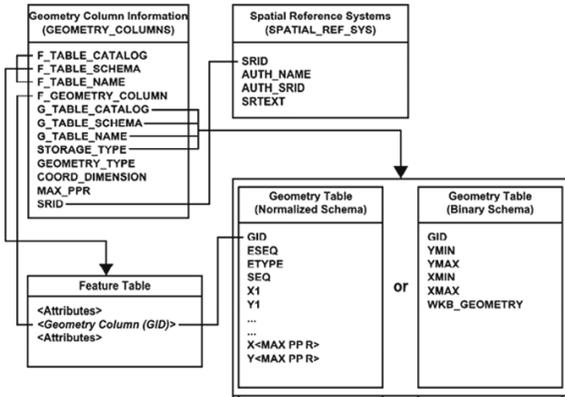


Fig: 1 Schema for feature tables using pre-defined data types

The GEOMETRY_COLUMNS table describes the available feature tables and their Geometry properties. GEOMETRY_COLUMNS table provides information on the feature table, spatial reference, geometry type, and coordinate dimension for each Geometry column in the database.

```
CREATE TABLE GEOMETRY_COLUMNS (
F_TABLE_CATALOG CHARACTER VARYING NOT
NULL,
F_TABLE_SCHEMA CHARACTER VARYING NOT
NULL,
F_TABLE_NAME CHARACTER VARYING NOT NULL,
F_GEOMETRY_COLUMN CHARACTER VARYING NOT
NULL,
.....
.....
.....
GEOMETRY_TYPE INTEGER,
COORD_DIMENSION INTEGER,
MAX_PPR INTEGER,
SRID INTEGER NOT NULL
REFERENCES SPATIAL_REF_SYS,
CONSTRAINT GC_PK PRIMARY KEY
(F_TABLE_CATALOG, F_TABLE_SCHEMA,
F_TABLE_NAME, F_GEOMETRY_COLUMN)
)
```

Example 1: Creation of Geometry Table

The SPATIAL_REF_SYS table describes the coordinate system and transformations for Geometry.

```
CREATE TABLE SPATIAL_REF_SYS
(
SRID INTEGER NOT NULL PRIMARY KEY,
AUTH_NAME CHARACTER VARYING,
```

```
AUTH_SRID INTEGER,
SRTEXT CHARACTER VARYING(2048)
)
```

Example 2: Creation of Spatial Reference System

The FEATURE TABLE stores a collection of features.

Metadata is required to store ancillary information about Geo database. Metadata is data about data. It allows common information to be stored at a common level. The attributes like owner name, version number, description of layer, remarks, annotation etc are stored in metadata.

```
CREATE TABLE ANNOTATION_TEXT_METADATA AS
{
F_TABLE_CATALOG AS CHARACTER VARYING NOT
NULL,
.....
.....
.....
```

```
A_TEXT_DEFAULT_MAP_BASE_SCALE AS CHARACTER
VARYING,
A_TEXT_DEFAULT_ATTRIBUTES AS CHARACTER
VARYING}
```

Example 3: Creation of Metadata

The geometry metadata describing the dimensions, lower and upper bounds and tolerance in each dimension is stored in a global table owned by MDSYS (which users should never directly update). Each Spatial user has the following views available in the schema associated with that user:

USER_SDO_GEOM_METADATA contains metadata information for all spatial tables owned by the user (schema). This is the only view that you can update, and it is the one in which Spatial users must insert metadata related to spatial tables. ALL_SDO_GEOM_METADATA contains meta-data information for all spatial tables on which the user has SELECT permission. Spatial users are responsible for populating these views. For each spatial column, you must insert an appropriate row into the USER_SDO_GEOM_METADATA view. Metadata prototype is given below.

```
INSERT INTO spatial_ref_sys VALUES
(101, 'POSC', 32214,
'PROJCS["UTM_ZONE_14N", GEOGCS["World Geodetic
System 72", DATUM["WGS_72", ELLIPSOID["NWL_10D",
6378135, 298.26]],
PRIMEM["Greenwich", 0], UNIT["Meter",1.0]],
PROJECTION["Transverse_Mercator"],
PARAMETER["False_Easting", 500000.0],
```

```
PARAMETER["False_Northing", 0.0],
PARAMETER["Central_Meridian", -99.0],
PARAMETER["Scale_Factor", 0.9996],
PARAMETER["Latitude_of_origin", 0.0],
UNIT["Meter", 1.0]]);
```

Example 4: Inserting Map Projection parameters

4 RATIONALE OF THE MODEL

This model is scalable to obtain certain objectives. The model implements centralization enterprise spatial data. If database is centralized, it will be more convenient for data manager to update, maintain and distribute. The model enables organizations to use the right tool for the job while eliminating complicated data transfers and multiple copies of the same data throughout the enterprise. The model also helps to implement data sharing between enterprises, especially as data sources for local emergency department. Because emergency department can access and maintain the all data they need. After all, data redundancy can be reduced.

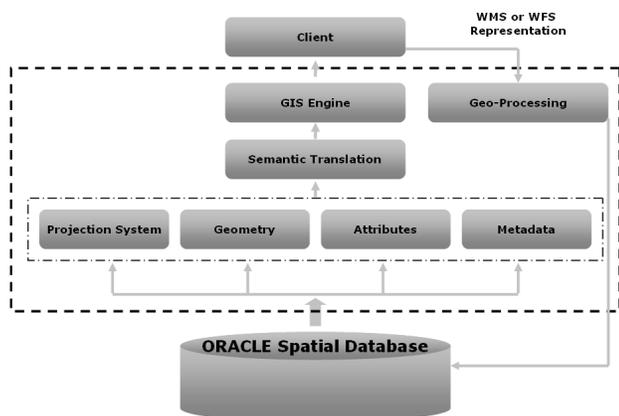


Fig: 2 Data translation component model

The oracle spatial data server can be established at centralized server. Different users can access Geo Spatial information through web browser. The application server is the main tier in internet model, which includes GIS application server, GIS application manager server and application connector. The application in this scenario is developed in Java using NetBeans 6.8 Interactive Development Environment (IDE).

5 GEO PROCESSING

Geo-processing is a GIS operation used to manipulate spatial data. A typical geo-processing operation takes an input dataset, performs an operation on that dataset, and returns the result of the operation as an output dataset. Common geo-

processing operations include geographic feature overlay, feature selection and analysis, topology processing, raster processing, and data conversion. Geo-processing allows for definition, management, and analysis of information used to form decisions. Gujarat state from India is used in this example.

Proposed System is developed in JAVA platform. JAVA SERVER PAGE (JSP) serves the solution in web interface. The Code snippet is given as below for Buffer Analysis using Geo Database.

CODE SNIPPET:

```
// ASSIGN VARIABLES
```

```
int sDragx=Integer.parseInt(startDragx);
int sDragy=Integer.parseInt(startDragy);
int eDragx=Integer.parseInt(endDragx);
int eDragy=Integer.parseInt(endDragy);
```

```
...
```

```
BufferedImage buffimage1=null;
JGeometry gmtry1,gmtry2;
STRUCT dbobj1,dbobj2;
Graphics2D g11,g22;
Shape sing_sh=null,sh_rect=null;
ResultSet rs1=null,rsbuff=null;
Statement st1=null;
Connection con=newconn();
```

```
//INITIALIZING MAP
```

```
st1=con.createStatement(rs1.TYPE_SCROLL_SENSITIVE,rs1
.CONCUR_UPDATABLE);
double
boundingbox[]=boundingBox("TABLE_GUJDISTRICT");
buffimage1 = new
BufferedImage(800,600,BufferedImage.TYPE_INT_RGB);
g11 = (Graphics2D)buffimage1.getGraphics();
g22 = (Graphics2D)g11;
```

```
g22.setRenderingHint(RenderingHints.KEY_ANTIALIASING,
RenderingHints.VALUE_ANTIALIAS_ON);
g22.setPaint(new Color(48,0,0));
g22.fillRect(0,0,800,600);
```

```
//FIND BOUND BOX VALUES OF A LAYER
```

```
double
boundingbox[]=boundingBox("TABLE_GUJDISTRICT");
```

```
//CALCULATE SCALE FACTOR
```

```
sx = clientwidth / mapw;
sy = clientheight / maph;
```

```
// EXECUTE BUFFER STATEMENT
rsbuff= st1.executeQuery("select
sdo_geom.sdo_buffer(mdys.sdo_geometry(2003,null,null,m
dsys.sdo_elem_info_array(1,1003,3),
mdsys.sdo_ordinate_array(" + Math.min(sDragx, eDragx) +
"," + Math.min(sDragy, eDragy) + "," + (Math.min(sDragx,
eDragx)+Math.abs(sDragx - eDragx) + "," +
(Math.min(sDragy, eDragy) +
Math.abs(sDragy - eDragy)) + ")), " + dist + ",0.005)
from " + currenttable);
```

```
//DRAW GEOMETRY ON MAP
dbobj2 = (STRUCT)rsbuff.getObject(1);
gmtry2 = JGeometry.load(dbobj2);
sing_sh=gmtry2.createShape();
g11.setColor(Color.blue);
g11.setStroke(new BasicStroke(1));
g11.draw(sing_sh);

dbobj1 = (STRUCT) rs1.getObject("GEOM");
gmtry1 = JGeometry.load(dbobj1);
JGeometry geomrect = new
JGeometry(2003,gmtry1.getSRID(),arrrect,ordrect);
sh_rect=geomrect.createShape();
```

Example 5: A prototype Code in Java Server Page to create Buffer using Geo database

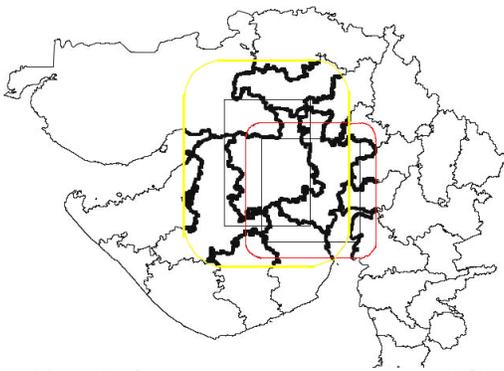


Fig 3(a): Buffer Creation for give Area of Interest (AOI)
Data source: BISAG

DNAME	DNAME	DNAME
AHMEDABAD	AHMEDABAD	AHMEDABAD
AMRELI	AMRELI	AMRELI
BANASKANTHA	ANAND	BANASKANTHA
BHAVNAGAR	BHARUCH	BHARUCH
JAMNAGAR	BHAVNAGAR	BHAVNAGAR
JUNAGADH	GANDHINAGAR	GANDHINAGAR
KACHCHH	KACHCHH	JAMNAGAR
MAHESANA	KHEDA	JUNAGADH
PATAN	MAHESANA	KACHCHH
RAJKOT	RAJKOT	KHEDA
SURENDRANAGAR	SURENDRANAGAR	MAHESANA
		PATAN
		RAJKOT
		SURENDRANAGAR

3 (b) Data

3 (c) Data

3 (d) Union of

inside Yellow Ring (buffered) inside Red Ring (buffered) Two Rings (buffered)

To perform geo-processing in the GIS system, it requires very efficient algorithm and good processing power.

It is required to find the association between two objects in Geo -processing. If two objects are A and B it can be associated like A touches B or A inside B or A equals B etc. Other GIS operations can be performed like difference between two geometries, Union, Intersection.

Difference Operation

Returns the region covered by A that is not covered by B. Function for this operation is SDO_DIFFERENCE



SDO_DIFFERENCE

XOR Operation

Returns the region of A and B that is not shared by both. This function is equivalent to $(A \cup B) - (A \cap B)$. Function for this operation is SDO_XOR



SDO_XOR

UNION Operation

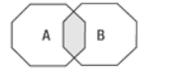
Returns the region of A and B that is covered by both. This function is equivalent to $(A \cup B)$



SDO_UNION

INTERSECTION Operation

Returns the region of A and B that is shared by both. This function is equivalent to $(A \cap B)$



SDO_INTERSECTION

Table 1: Various operations based on Set theory

6 TOPOLOGICAL RELATIONSHIPS

Topology is a major area of mathematics concerned with properties that are preserved under continuous deformations of objects, such as deformations that involve stretching, but no tearing or gluing. It emerged through the development of concepts from geometry and set theory, such as space, dimension, and transformation.

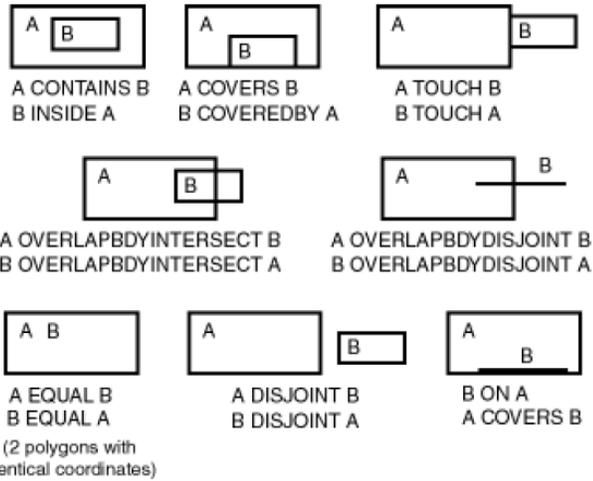


Fig 4 Possible association combinations between objects A and B

7 METHODOLOGY

Use of Oracle Geo Database can be utilized by converging Information Communication Technology (ICT). The Geo Database communicates with Web based GIS Map engine. The GIS map engine talks with client with authentication from Java Application Programming Interface (API).

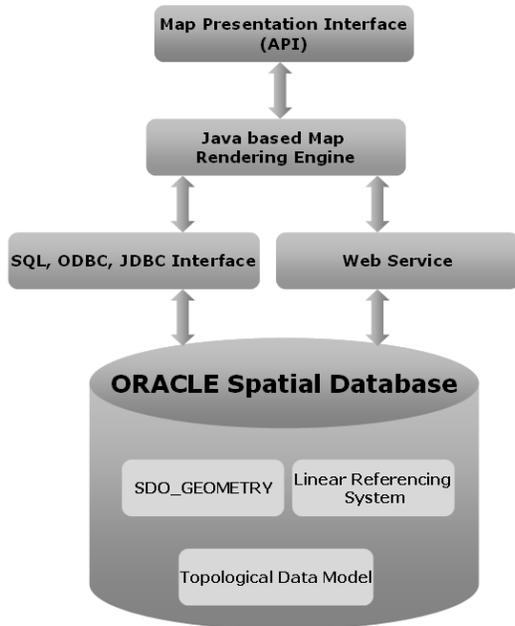


Fig: 5 Interface to access Oracle Spatial Geometry

8 3D RENDERING

Oracle spatial can extrude a Two-Dimensional Geometry to Three Dimension. Two-dimensional footprints of buildings can be extruded using the EXTRUDE function in the

SDO_UTIL package to erect a building on the two-dimensional footprint by specifying the ground height and the top height for each vertex of the two-dimensional geometry.

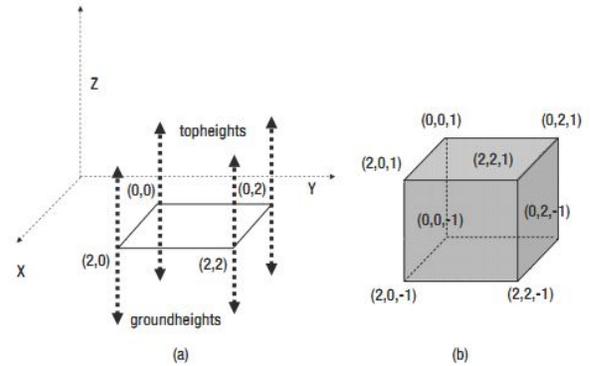


Fig 6(a) Example of a two-dimensional solid with the top heights and ground heights specified and Fig 6(b) the extruded solid object

```
SDO_UTIL.EXTRUDE
(
  geometry           SDO_GEOMETRY,
  groundheights     SDO_NUMBER_ARRAY,
  topheights        SDO_NUMBER_ARRAY,
  result_to_be_validated VARCHAR2,
  tolerance         NUMBER
) RETURN SDO_GEOMETRY
```

Example 6: Extrude schema

The parameters can be interpreted as shown below:

geometry: This specifies the input two-dimensional SDO_GEOMETRY object that needs to be extruded.

groundheights: This is an array of numbers, one each for each vertex for use as the ground height (minimum z value). If only one number is specified, then all vertices get the same value (that is specified here).

topheights: This is an array of numbers, one each for each vertex for use as the top height (maximum z value). If only one number is specified, then all vertices get the same value (that is specified here).

result_to_be_validated: This is a character string that can be set to either 'TRUE' or 'FALSE'. This string informs Oracle whether to validate the resulting geometry.

tolerance: This specifies the tolerance to use to validate the geometry.

A prototype example extruding a Polygon to a Three-Dimensional Solid is given below.

```
SELECT SDO_UTIL.EXTRUDE(
  SDO_GEOMETRY -- first argument to validate is
  geometry
  (
```

```

2003, -- 2-D Polygon
NULL,
NULL,
SDO_ELEM_INFO_ARRAY(1, 1003, 1 -- A polygon element
),
SDO_ORDINATE_ARRAY (0,0, 2,0, 2,2, 0,2, 0,0) -- vertices
of polygon
),
SDO_NUMBER_ARRAY(-1), -- Just 1 ground height value
applied to all vertices
SDO_NUMBER_ARRAY(1), -- Just 1 top height value
applied to all vertices
'FALSE', -- No need to validate
0.5 -- Tolerance value
) EXTRUDED_GEOM FROM DUAL;

```

Example 7: Extruding a sample data using SQL statement

```

EXTRUDED_GEOM(SDO_GTYPE, SDO_SRID,
SDO_POINT(X, Y, Z), SDO_ELEM_INFO, SDO_ORDINA
-----
SDO_GEOMETRY(
3008, -- 3-Dimensional Solid Type
NULL, NULL,
SDO_ELEM_INFO_ARRAY(
1, 1007, 1, -- Solid Element
1, 1006, 6, -- 1 Outer Composite Surface made up of 6
polygons
1, 1003, 1, -- First polygon element starts at offset 1 in
SDO_ORDINATES array
16, 1003, 1, -- second polygon element starts at offset 16
31, 1003, 1, -- third polygon element starts at offset 31
46, 1003, 1, -- fourth polygon element starts at offset 46
61, 1003, 1, -- fifth polygon element starts at offset 61
76, 1003, 1, -- sixth polygon element starts at offset 76
SDO_ORDINATE_ARRAY( -- ordinates storing the vertices
of the polygons
0, 0, -1, 0, 2, -1, 2, 2, -1, 2, 0, -1, 0, 0, -1,
0, 0, 1, 2, 0, 1, 2, 2, 1, 0, 2, 1, 0, 0, 1, 0, 0,
-1, 2, 0, -1, 2, 0, 1, 0, 0, 1, 0, 0, -1, 2, 0,
-1, 2, 2, -1, 2, 2, 1, 2, 0, 1, 2, 0, -1, 2, 2,
-1, 0, 2, -1, 0, 2, 1, 2, 2, 1, 2, 2, -1, 0, 2,
-1, 0, 0, -1, 0, 0, 1, 0, 2, 1, 0, 2, -1))

```

Example 8: output of Extruded sample

Following example shows the satellite image of a building in New York City, USA taken from Google map. The Building ground height and top height are measured in the form of latitude, longitude and altitude with Global Positioning System (GPS). The building corner points are marked and respective co-ordinates are extruded using SDO_UTIL.EXTRUDE method.



Fig: 7(a) Building in a satellite image with edges marked with dark spots.

The extruded geometry vertices returns an array that is plotted as an object as shown in the below figure.

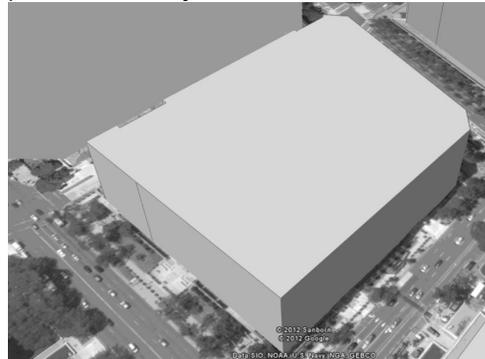


Fig 7(b): Extruded building

9 APPLICATIONS

The proposed method is used to derive new datasets from the data warehouse. The geo-processing kind of tasks that require very efficient algorithms can be effectively carried out by leveraging the potential of Oracle spatial geo database. 3D Elevation Model, Triangulation Irregular Network (TIN) can also be generated from geo database. Moreover, centralized data-centric and secured applications can be developed and scaled at enterprise level.

10 LIMITATIONS

The geo-processing task and extrude task are memory intensive. The proposed methodology can be optimized by introducing cloud based storage and routing mechanism. Parallel Processing can also optimize the performance to render the output in terms of geographical maps.

11 CONCLUSIONS

The proposed solution is scalable to enterprise level. The Spatial Data Infrastructure (SDI) can be established that can be helpful to derive spatial data and attributes from various datasets of different organizations in multiple combinations that can provide near real time information resulting in

efficient and effective decision support system (DSS) to help multiple areas like Health, Defense, Disaster etc.

ACKNOWLEDGEMENTS

The authors would like to thank to the Director, BISAG, T .P. Singh for his inspiration and motivation.

REFERENCES

- [1] Oracle Spatial User guide and Reference Release 2002, a96630.pdf
- [2] OpenGIS Coordinate Transformation Service Implementation Specification from <http://www.opengeospatial.org/standards>.
- [3] <http://www.opengeospatial.org/ogc>
- [4] <http://en.wikipedia.org/wiki/Geoprocessing>
- [5] <http://en.wikipedia.org/wiki/Topology>
- [6] Z.L. Li, R.L. Zhao, and J. Chen, "A generic algebra for spatial relations," Progressing in Natural Science, vol. 12, no. 7, pp. 528-536, July 2002.Oracle® Spatial User's Guide and Reference Release 9.2
- [7] Ravi Kothuri, Albert Godfrind, and Euro Beinat, Pro Oracle® Spatial for Oracle Database 11g, Apress, 2007
- [8] C. Murray, "Oracle spatial Topology and Network Data Models," Oracle Corporation, pp. 193-251, 2005
- [9] http://wikihelp.autodesk.com/Infr._Map_Server/enu/2012/Help/0000-Administ0/0104-Data_Mod104/0182-Data_Mod182

BER Performance of Linear Multi-user Detectors in DS-CDMA

D.Swathi ¹, J.Ravindrababu ², B. Alekya ³, Dr.E.V.Krishna Rao⁴

Abstract: - This paper examines the Bit Error Rate (BER) performance of Linear Multi-user Detectors in Direct Sequence Code Division Multiple Access (DS-CDMA) system. Multiple access interference (MAI) limits the capacity of Direct Sequence Code Division Multiple Access (DS-CDMA) systems. In CDMA systems MAI is considered as additive noise and a matched filter bank is employed. Multi-user detectors are classified as optimal and suboptimal. The main drawback of the optimal multi-user detection is complexity so that suboptimal approaches are being sought. Much of the present research is aimed at finding an appropriate tradeoff between complexity and performance. These suboptimal techniques have linear and non-linear algorithms. In this paper, introduce linear Multi-user Detectors in Direct Sequence Code Division Multiple Access (DS-CDMA) system. Analysis is to be carried out and simulations to be done.

Keywords: DS-CDMA, MF, Decorrelator, MMSE, ZF .MAI, Gold sequence.

I. INTRODUCTION

The Capacity of Frequency Division Multiple Access (FDMA) or Time Division Multiple Access (TDMA) or hybrids, common in the 2nd generation, is well defined when RF channels or time slots are no longer available no more customers can be accommodated. It is possible to include more users, although at the price of a slightly worse signal-to-interference ratio for everyone.

In DS-CDMA communication system, users are multiplexed by distinct codes rather than by orthogonal frequency bands or by orthogonal time slots. A conventional DS-CDMA detector follows a single user detection strategy in which each user is filter just treat the MAI as additive white Gaussian noise (AWGN). However, unlike AWGN, MAI has a nice correlative structure that is quantified treated separately as a signal, while the other users are considered as either interference or noise. Multi-user detection is a technology that spawned in the early 80's. It has now developed into an important, full-fledged field in multi-access communications. Multi-user Detection (MUD) is the intelligent estimation / demodulation of transmitted bits in the presence of Multiple Access Interference (MAI). MAI occurs in multi-access communication systems (CDMA/ TDMA/FDMA) where simultaneously occurring digital streams of information interfere with each other. Conventional detectors based on the matched by the cross-correlation matrix of the signature sequences. Hence, detectors that take into account this correlation would perform better than the conventional filter-bank [1-7].

- D.Swathi is currently pursuing masters degree program in microwav communication engineering in p.v.p.s.i.t j.n.t.university, kakinada, In mail: dswathi289@gmail.com
- J.Ravindrababu is currently working as Associate Professor , departm electronics and communication engineering, P.V.P.Siddhartha Institu Technology, j.n.t.university, Kakinada, India E-mail: jrb0009@gmail.com.

II. SYSTEM MODEL

MUD is basically the design of signal processing algorithms that run in the black box shown in figure 1. These algorithms take into account the correlative structure of the MAI. The K-user discrete time basic synchronous CDMA model has been used throughout the development of this paper. The case of antipodally modulated user information (BPSK modulation) spread using BPSK

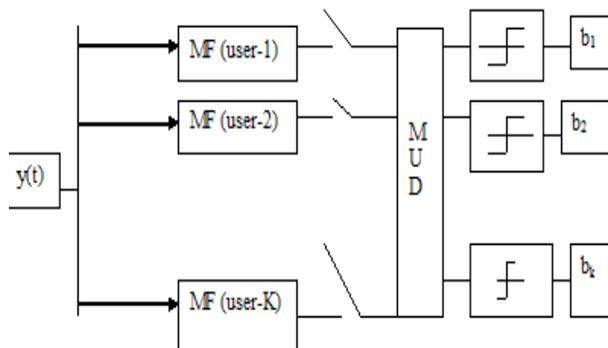


Figure.1 A typical multi-user detector

The signal at the receiver is given by

$$y(t) = \sum_{k=1}^K A_k B_k S_k(t) + n(t) \text{ --- (1)}$$

Where

S_k is the signature waveform of the k^{th} user (S_k is ormalized to have unit energy) i.e.,

$$\langle S_k, S_k \rangle = 1$$

Where

A_k is the received amplitude of the k^{th} user

- B_k is the input bit of the k^{th} user, $b_k \in \{-1, 1\}$.
- $n(t)$ is additive white Gaussian noise with PSD N_0 .

Since synchronous CDMA is considered, it is assumed that the receiver has some means of achieving perfect chip synchronization.

The cross-correlation of the signature sequences are defined as

$$\rho_{ij} = \langle S_i S_j \rangle = \sum_{k=1}^N S_i(k) S_j(k) \quad \text{--- (2)}$$

Where N is the length of the signature sequence
 The cross-correlation matrix is then defined as

$$R = \rho_{ij}$$

R is a symmetric, non-negative definite, toeplitz matrix

III. MATCHED -FILTER

Introduces and analyses the matched filter bank detector which was the conventional and simplest way of demodulating CDMA signals (or any other set of mutually interfering digital streams). The matched filter also forms the front-end in most MUDs and hence understanding the operation is crucial in appreciating the evolution of MUD Technology. In conventional single-user digital communication systems, the matched filter is used to generate sufficient statistics for signal detection. In the case of a multi-user system, the detector consists of a bank of matched filters (Each matched to the signature waveforms of different users in the case of CDMA). This is shown in figure 2. This type of detector is referred to as the conventional detector in MUD literature. It is worth mentioning that we need exact knowledge of the users signature sequences and the signal timing in order to implement this detector [8].

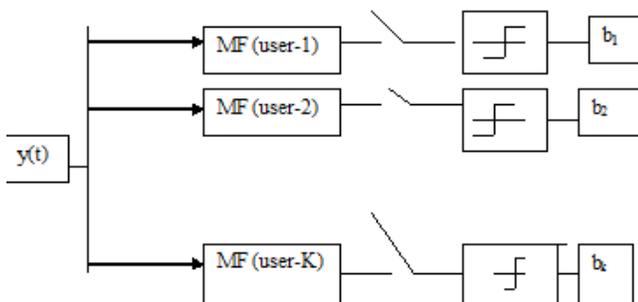


Figure 2 A matched filter bank

The decision statistic the output of the K^{th} matched filter is given by

$$y_k = \int_0^T y(t) s_k(t) dt \quad \text{--- (3)}$$

Expanding this equation

$$y_k = \int_0^T \left\{ \sum_{j=1}^K A_j B_j S_j(t) + n(t) \right\} S_k(t) dt \quad \text{--- (4)}$$

Therefore

$$y = RAb + n \quad \text{--- (5)}$$

IV. DECORRELATING DETECTOR

An optimal receiver must be capable of decoding the bits error-free when the noise power is zero. The decorrelating detector is investigated. This detector makes use of the structure of MAI to improve the performance of the matched filter bank. The decorrelating detector falls into the category of linear multi-user detectors. As shown in figure 3, the decorrelating detector operates by processing the output of the matched filter bank with the R^{-1} operator where R is the cross-correlation matrix.

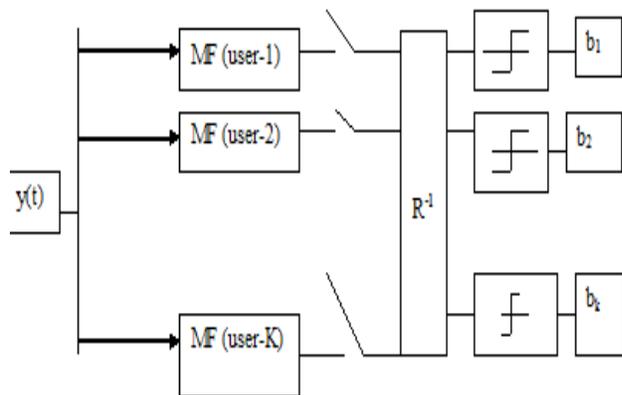


Figure 3. Decorrelating Detector

$$\hat{b} = \text{sgn}(R^{-1}(RAb + n)) \quad \text{--- (6)}$$

$$\hat{b} = \text{sgn}(Ab + R^{-1}n) \quad \text{--- (7)}$$

Hence, we observe that in the absence of background noise the decorrelating detector achieves perfect demodulation unlike the matched filter bank. One advantage of the decorrelating detector is that it does not require knowledge of the received signal amplitudes. The

decorrelating receiver performs only linear operations on the received statistic and hence it is indeed a linear detector. The decorrelating detector is proved to be optimal under 3 different criteria: least squares, near-far resistance and maximum-likelihood [8].

V MMSE LINEAR DETECTOR

The MMSE receiver is another kind of linear multi-user receivers. The description of MMSE detector can be graphically represented in Figure 4. The MMSE implements the linear mapping which minimizes the mean-squared error between the actual data and the soft output of the conventional detector, so the decision for the k^{th} user is made based on in this approach where the mean squared error between the output and data is minimized. The detector resulting from the MMSE (minimum mean square error) criterion is a linear detector.

$$\hat{b} = (R + N_0 A^2)^{-1} \dots (8)$$

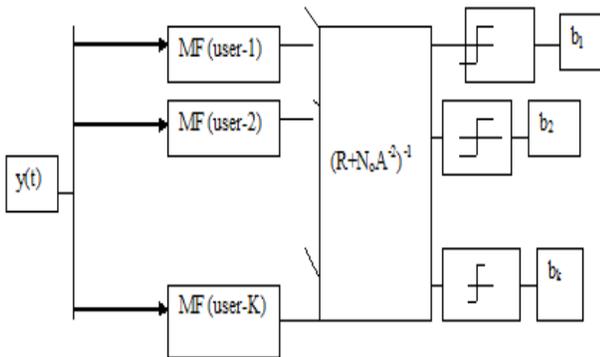


Figure 4 MMSE linear detector

VI. ZERO-FORCING DETECTOR

The zero-forcing receiver is a natural progression of the decorrelating detector. Now that we have removed the MAI, we want to eliminate the ISI as well. This can be done by taking into consideration each users channel impulse response. The zero forcing equalizer is successful at eliminating MAI and ISI, but has some tradeoffs. Also, the zero-forcing equalizer suffers the noise enhancement problems as does the decorrelating detector. But In order to improved performance in the zero forcing detector in presence of noise[9].

VII. SIMULATION RESULTS

Figure A, B, C and D show the error rate performance of the bank of matched filter. Decorralator , MMSE and ZF. The simulation scenario is observed that as the MAI increases (the number of users increases) the performance becomes

poor. But the decorralator is better performed than MF. Similarly the MMSE is better performed than decorralator and matched filter. Similarly like this the zero forcing detector is also well performed compared to other detectors.

Figure E, F, and G shows the comparison of error performance of different detectors. The zero forcing detector is well performed compared to the other detector in all cases like 2-user, 5-user and also 10-user case .

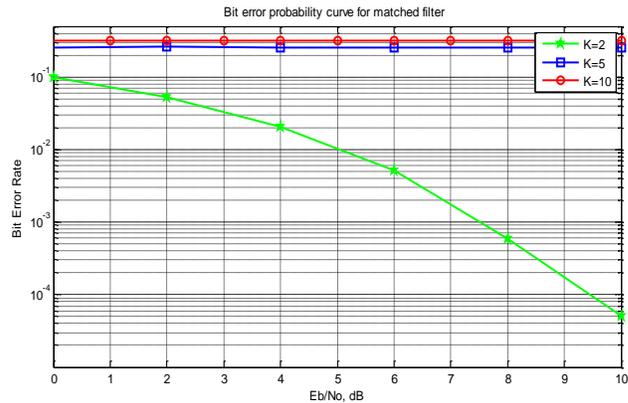


Figure-A: performance of Matched filter

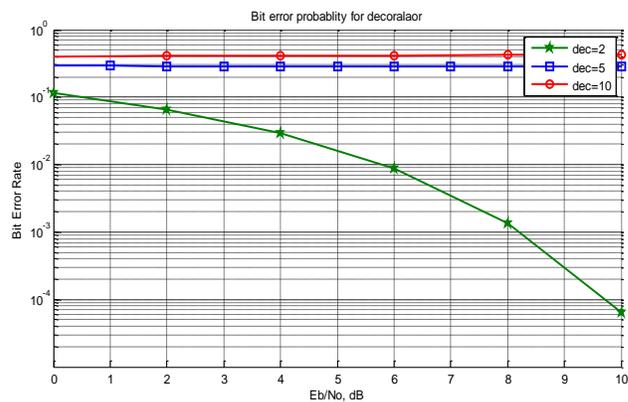


Figure-B: performance of Decoralator

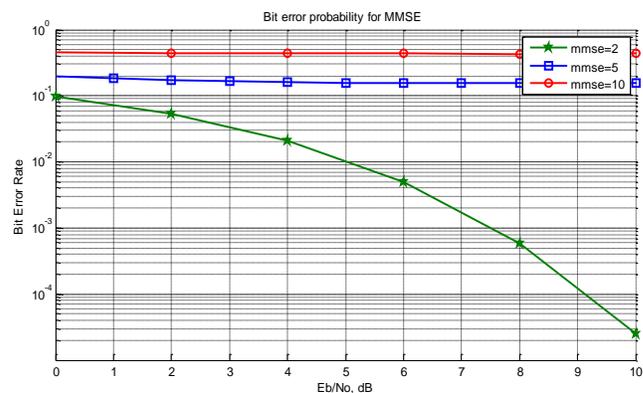


Figure-C: performance of MMSE

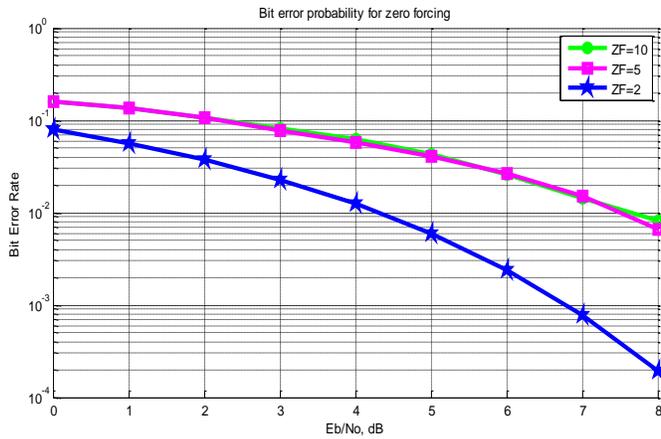


Figure-D: performance of ZF

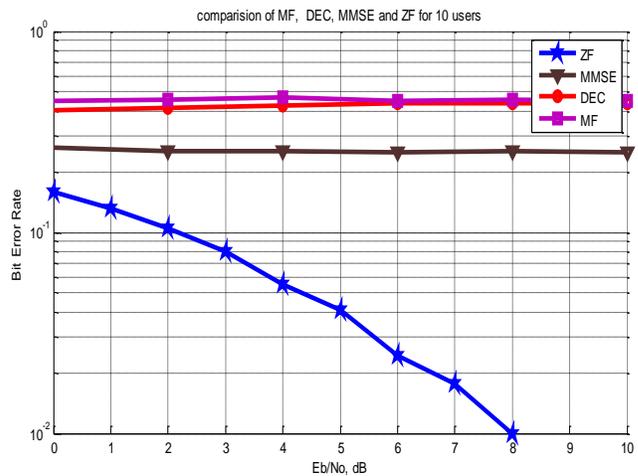


Figure-G: Comparison of Detectors for 10-user

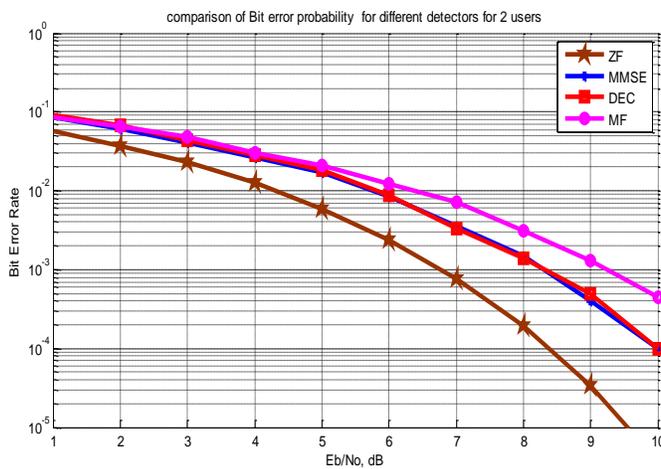


Figure-E: Comparison of Detectors for 2-user

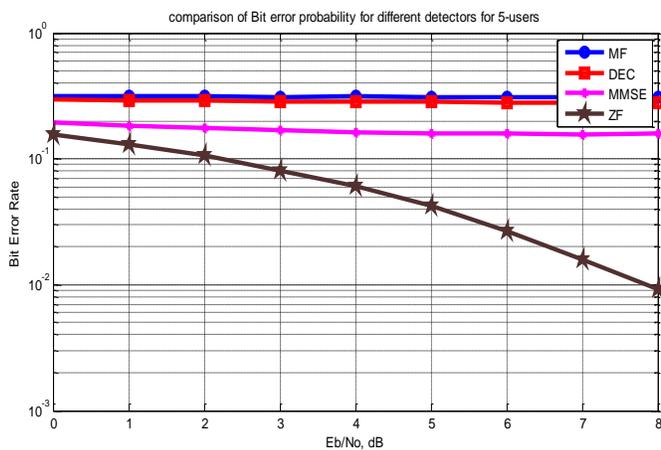


Figure-F: Comparison of Detectors for 5-user

VIII. CONCLUSIONS

This Paper is a compilation of different approaches to linear multi-user detection. The requirement of this technology was motivated by studying the conventional detector. The matched filter bank just ignores the correlative structure of the MAI present in CDMA systems. Further, it was also shown that in the absence of noise, the conventional detector is a totally unreliable detector. This called for the need for better detectors. The decorrelating detector was then introduced which takes the conventional detector one step further by incorporating the correlative structure of the MAI in the detection. This implied that the decorrelating detector could be improved upon. The MMSE linear detector was then shown to take the decorrelating detector one step further by incorporating some SNR information along with the correlative structure of MAI. Thus, the performance was better than the decorrelating detector at high SNRs. It must also be noted that when the background noise is totally absent (infinite SNR). Finally the zero forcing detector is well performed. The choice of the MUD algorithm depends on a lot of factors like the application, channel information available, availability of training sequences, complexity cost and overhead involved.

IX. References

- [1] A. J. Viterbi (1994), The Orthogonal-Random Waveform Dichotomy for Digital Mobile Personal Communications, *IEEE Pers. Commun*, 1st qtr., pp. 18-24.
- [2] C. Kchao and G. Stuber (1993), Performance Analysis of a Single Cell Direct Sequence Mobile Radio System, *IEEE Trans. on Commun.*, vol. COM-41, no 10, pp. 1507-1516.
- [3] Clark, George C., Jr., and J. Bibb Cain (1981), *Error-Correction Coding for Digital Communications*, New York: Plenum Press.
- [4] D. V. Sarwate and M. B. Pursley (1980), Cross correlation Properties of Pseudorandom and Related Sequences, *Proc. IEEE*, vol. 68, no. 5, pp. 593-619.
- [5] J. C. Liberti (1996), *Spatial Processing for Higher Wireless Systems*, Bellcore Pub. IM-558.
- [6] J.G. Proakis (1995), *Digital Communications*, 3rd Edition, New York: McGraw-Hill.
- [7] Jochen Schiller (2003), *Mobile Communications*, 2nd Edition, Addison- Wesley.
- [8] S. Verdu, *Multiuser Detection* Cambridge University Press, 1998.
- [9] D.R. Brown, D.L. Anair, C.R. Johnson, Jr., "Linear Detector Length Conditions for DS-CDMA Perfect Symbol Recovery" to appear in *Proc. IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, Mar. 1999.

DSS for Library Management System Implementing RFID

Sanjay Waykar, Amruta Shelar, Sushant Zanjure, Sayali Vibhute, Abhishek Singh

Abstract -- Radio frequency based identification (RFID) based library management using decision support system(DSS) is the most rare and uniquely found integration of two cutting edge technologies that is RFID and DSS, which has led to tremendous increase in the speed, efficiency and accuracy of the library. RFID is a new generation of Auto Identification and Data collection technology which helps to automate business processes and allows identification of large number of tagged objects like books using radio waves. This system is based on UHF RFID readers, supported with antennas at gate and transaction sections, and library cards containing RFID-transponders which are able to electronically store information that can be read / written even without the physical contact with the help of radio medium RFID would allow fast transaction flow for the library and will prove immediate and long term benefits to library in traceability and security. In addition to it, we are providing decision support system for library to make it easier for customer to decide and locate their requirements without manual intervention. DSS is implemented using various data mining techniques for arranging the dispersed information of books, customer, administrators etc in a systematic manner. This unique combination of RFID and DSS is expected to enhance the services offered by the library to a great extent.

Index Terms – Decision Support System (DSS), Radio Frequency Identification (RFID), Data Mining, Library Management System, Barcode.

◆

1 INTRODUCTION

A library is a huge collection of information, sources, resources, books, and services, and the structure in which it is housed. Due to financial and human constraints for library support, library managers increasingly encounter the need to justify everything they do, for example, the library's operation budget in particular. The most frustrating problem they face is knowing that the information needed is available somewhere in the ocean of data but there is no easy way to obtain it. For example, it is not easy to ascertain whether the materials of a certain subject area, which consumed a lot of financial resources for their acquisition and processing, are either frequently used (i.e., a high rate of circulation), seldom - used or not used at all, or whether they satisfy users' needs.

DSS for library management system implementing RFID is expected to solve such problems faced by the librarian, as well as the library users. The unique feature of this proposed system is the combination of RFID and DSS. Till date, DSS and RFID have been implemented in various libraries, but separately. The proposed system attempts to integrate them, which will prove beneficial to the library. Both these technologies have their own individual benefits. Thus, integration of RFID and DSS will cause combination of features owned by both by technologies imparting it higher level of efficiency.

This library management system replaces the traditional, manual book keeping of records. It intends to replace the barcodes by RFID tags. The RFID based LMS facilitates the fast issuing, reissuing and returning of books with the help of RFID enabled modules. It directly provides

the book information and library member information to the library management system and does not need the manual typing. The concept of DSS, which includes Data Mining will help the librarian to provide many services, such as managing the shelf according to the interest of the people. The general description of the technologies our system is based upon is stated as follows:-

1.1 RFID

The "Radio Frequency Identification (RFID) is an automatic identification system. RFID uses RF to identify "tagged" items. This data is then collected and transmitted to a host system using an RF Reader. The data transmitted by the tag may provide identification or location information, or specifics about the product tagged, such as price, colour, date of purchase, etc."

1.1.1 TAGS USED IN RFID

Tags are typically composed of a microchip for storage and computation, and a coupling element, such as an antenna coil for communication. Tags may also contain a contact pad. Tag memory may be read-only, write-once read-many or fully rewritable.

Broadly the tags have been classified in three categories:

- Active Tag: An active RFID tag is equipped with a power source for the tag's circuitry and antenna. The advantages of an active RFID tag includes readability from a distances of one hundred feet or more as well as capability to have other sensors that can use electricity for power. The major disadvantages of an active RFID tag are the limitations on

the lifetime of the tag (5 years). They are more expensive and physically larger and they add to the maintenance cost if the batteries are replaced. Battery outages in an active tag can result in expensive misreads.

- **Passive Tag:** Passive RFID tag does not contain a power source; the power is supplied by the reader. The tag draws power from the inductive coupling with reader antenna. The major disadvantages of a passive tag are that the tag can be read only at very short distances, typically a few feet at most. However there are many advantages. The tag functions without a battery which increases the life time to more than 20 years. The tags are less expensive and much smaller. These tags have almost unlimited applications in consumer goods and other areas.

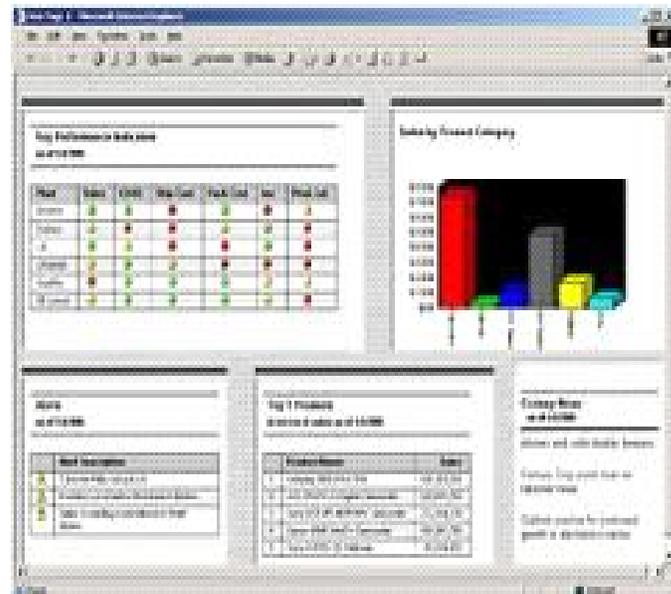
- **Semi-Passive Tag:** Like passive tags, semi-passive tags reflect (rather than transmit) RF energy back to the tag reader to send identification information. However, these tags also contain a battery that powers their ICs. This allows for some interesting applications, such as when a sensor is included in the tag so it can transmit real-time attributes, such as temperature, humidity, and timestamp. By using the battery only to power a simple IC and sensor—and not including a transmitter—the semi-passive tags achieves a compromise between cost, size, and range.

In the proposed system, use of RFID is mainly for identifying books or individual users, administrators using smartcards and tags. Preferably, passive tags will be used here, as tags are required only for reading within short distances or range.

Another most important concept we are going to implement in our system is Decision Support System (DSS).

1.2 DECISION SUPPORT SYSTEM

The best decision support system provides high-level summaries and drilldowns to details. Decision Support Systems (DSS) are a specific class of computerized information system that supports business and organizational decision-making activities. A properly designed DSS is an interactive software-based system intended to help decision makers compile useful information from raw data, documents, personal knowledge, and/or business models to identify and solve problems and make decisions.



Decision support systems could be categorized in terms of the generic operations that can be performed by systems. These generic operations extend along a single dimension, ranging from extremely data-oriented to extremely model-oriented. Alter conducted a field study of 56 DSS that he categorized into seven distinct types of DSS. The seven types include:

- File drawer systems that provide access to data items.
- Data analysis systems that support the manipulation of data by computerized tools tailored to a specific task and setting or by more general tools and operators.
- Analysis information systems that provide access to a series of decision-oriented databases and small models.
- Accounting and financial models that calculate the consequences of possible actions.
- Representational models that estimate the consequences of actions on the basis of simulation models.
- Optimization models that provide guidelines for action by generating an optimal solution consistent with a series of constraints.
- Suggestion models that perform the logical processing leading to a specific suggested decision for a fairly structured or well-understood task.

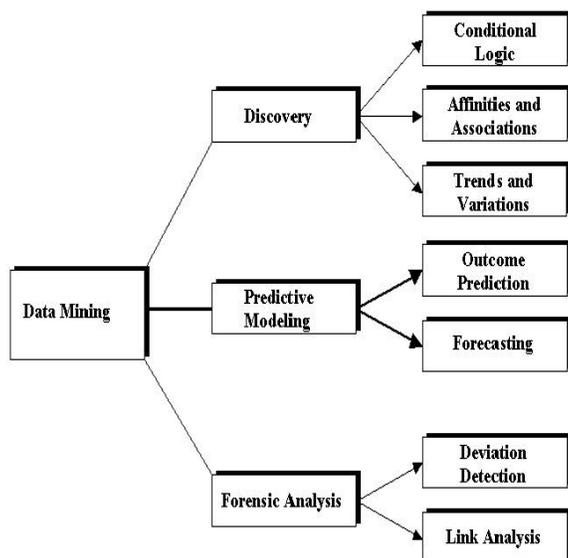
We are going to use DSS for general activities in library such as shelf management, stock management, searching of books.

An important part of DSS is data mining

4 DATA MINING

IBM defines Data Mining as, “the process of extracting previously unknown, valid and actionable information from large databases and then using the information to make crucial business decisions”

4.1 Classes of data mining



First, the information is previously unknown, in that it is not directly derived from the data. Instead, the information takes the form of relationships amongst the database columns where the value in one or more columns predicts the outcome in another – hence the name, predictive model. But predictive models must be valid. Rating a model's predictive power usually involves testing it against another data set. Finally, it goes without saying that the information is actionable. But this does suggest that there should be some goal in mind before putting any effort into data mining.

In our system, data mining will attempt to provide library patrons with more choice, and thus increase user satisfaction.

2 RELATED TECHNOLOGY

2.1 BARCODE USED IN LIBRARY SYSTEM

In this Approach of library Management System, traditional manual method of book keeping is reflected by automated library management system. Application of this concept is Surpass [12]. It's software for libraries which was developed back in 1982 for automated library management system. Surpass is a full-featured suite of applications that smoothly automates all of the day-to-day functions of libraries. Surpass is an effective automation solution for primary and secondary school districts as well as for public, college, corporate, and specialty libraries. Surpass offers a special package for church libraries too. Surpass Hosting Service is offers libraries the ability to put their whole system on-line with no need for in-house servers or IT staff.

- Barcode Scanning

- Catalog Management
- Circulation Management
- Custom User Interface
- Customer Database
- Customizable Fields
- Customizable Functionality
- Customizable Queries
- Customizable Reporting
- Data Import/Export
- Fee Collection
- Kiosk Integration
- OPAC
- Periodicals Management
- Reporting
- Search Functionality
- Self Service Features

Available on the following platforms: Windows, Web Based Support available: Regular Business Hours
Comparing our system with Surpass the one key advantage is replacement of RFID with Barcode.

2.2 RFID VERSUS BARCODE

- Barcode readers require a direct line of sight to the printed barcode; RFID readers do not require a direct line of sight to either Active RFID or passive RFID tags.
- RFID tags can be read at much greater distances; an RFID reader can pull information from a tag at distances up to 300 feet. The range to read a barcode is much less, typically no more than fifteen feet.
- RFID readers can interrogate, or read, RFID tags much faster; read rates of forty or more tags per second are possible. Reading barcodes is much more time-consuming; due to the fact that a direct line of sight is required, if the items are not properly oriented to the reader it may take seconds to read an individual tag. Barcode readers usually take a half-second or more to successfully complete a read.
- Line of sight requirements also limit the ruggedness of barcodes as well as the reusability of barcodes. (Since line of sight is required for barcodes, the printed barcode must be exposed on the outside of the product, where it is subject to greater wear and tear.) RFID tags are typically more rugged, since the electronic components are better protected in a plastic cover. RFID tags can also be implanted within the product itself, guaranteeing greater ruggedness and reusability.
- Barcodes have no read/write capability; that is, you cannot add to the information written on a printed barcode. RFID tags, however, can be read/write devices; the RFID reader can communicate with the tag, and alter as much of the information as the tag design will allow.
- RFID tags are typically more expensive than barcodes, in some cases, much more so.

2.3 LIBBEST

LibBest [14], Library RFID Management System. LibBest is the software provider who provides RFID based Library Management System. It provides the following functionalities.

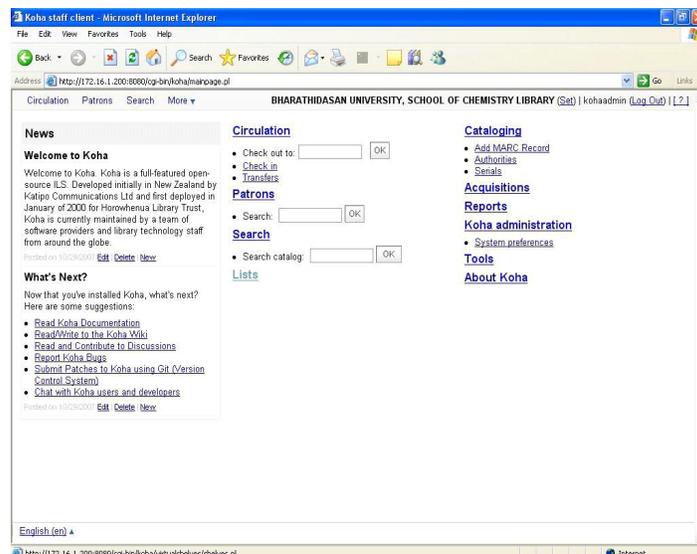
- RFID tags replace both the EM security strips and Barcode.
- Simplify patron self check-out / check-in.
- Ability to handle material without exception for video and audio tapes.
- Radio Frequency anti-theft detection is innovative and safe.
- High-speed inventory and identify items which are out of proper order.
- Long-term development guarantee when using Open Standard.

This system leverages the benefit of RFID for faster processing. Comparing our system with this LibBest system, we are going to provide Decision Support for the user of the system and the administrator in addition to the benefits of the RFID technology.

2.4 EXISTING LIBRARY SYSTEMS

Traditional Library Systems consisted of manual system of managing various activities. The system consisted of large volume of books and can maintain student's data in form of registers that consisted of all details of student members. Issue procedure for the books to issue books allotted to each student. The librarian takes off the book card containing all details of the book that is kept in the book and attaches with the issue card of the student and keeps in record column. Book return is inverse process of issue book when the student returns back the book. Also there was provision of requesting new book for the library by filling the new book request form and deletion of highly used and damaged was done by cancelling the record of the book from all the registers. Fine system required, librarian to calculate the fine for the student which is paid by the student on not returning the book within return date of the book. Also no backup was maintained and duplication of data consumes lot of time.

Here, from the sources [3] we get acquainted that Library automation refers to mechanization of library housekeeping operations predominantly by computerization. Systems have been developed to update the database of Books and other Resources of the university, to implement automated system using respective Library Integrated Open Source Software, to carry out the charging and discharging functions of the circulation section more effectively, to provide various search options, to know the availability of books in the Library, to generate the list of books due by a particular member and also the overdue charges.



The Greenstone Digital Library Software [4] provided a way of building and distributing digital library collections, opening up new possibilities for organizing information and making it available over the Internet or on CD-ROM. Produced by the New Zealand Digital Library project [5], Greenstone was intended to lower the bar for construction of practical digital libraries, yet at the same time leave a great deal of flexibility in the hands of the user.

Greenstone has been used to make many digital library collections. Some were created within the New Zealand Digital Library as demonstration collections. However, the use of Greenstone internationally is growing rapidly, and several web sites show collections created by external users. Most contain unusual and interesting material, presented in novel and imaginative ways. This article [5] briefly reviewed a selection of Greenstone digital library sites to give a feeling for how Greenstone is being used for public digital libraries throughout the world from different countries, from different kinds of library, with different sorts of source material.

The research and surveys [6] also revealed that there was development of online library management projects too. That kept track of issues and submission of books and also created online memberships. It provided better and efficient service to the members and reduced workload of employees with fast retrieval of information about the books and that too with just a click. There were also projects that were offline providing facilities like the books received in the library are entered in Books Entry form and the new student is entered in the student entry form. When the student wants to get the desired book the same is issued on the availability basis to the student. The issuance and due date for the returning of the book is also entered into the Book Issue form under third menu Book Issue. It used Visual Basic which is Graphical Rapid Application Development (RAD) tool that aims at

providing the user with a graphical interface that is intuitive and easy to use.

The research reveals that libraries began using RFID systems to replace the electromagnetic and barcodes in late 1990s. Approximately 130 libraries in library in North America are using RFID but hundreds more started considering it. RFID is a technology that is sparking interest in the library community because of its applications that promise to increase efficiency, productivity and enhance user satisfaction. In each and every activity within a library, RFID technology will provide a greater amount of efficiency and error free functioning. It will ensure that there is:

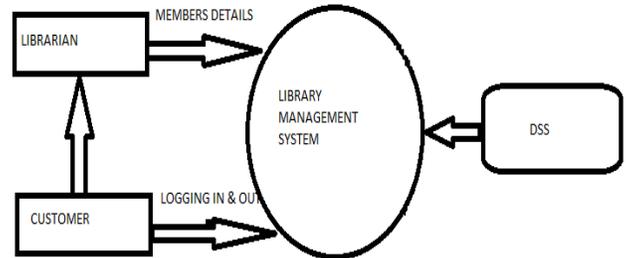
- Quick check-in/check-out of items for convenience of members
- Quick and correct shelving of items
- Complete prevention of thefts
- Quick inventory check

RFID can be used in library circulation operations and theft detection systems. RFID-based systems move beyond security to become tracking systems that combine security with more efficient tracking of materials throughout the library, including easier and faster charge and discharge, inventorying, and materials handling (Boss 2004). This technology helps librarians reduce valuable staff time spent scanning barcodes while charging and discharging items. RFID is a combination of radio - frequency-based technology and microchip technology. The information contained on microchips in the tags affixed to library materials is read using radio frequency technology, regardless of item orientation or alignment (i.e., the technology does not require line-of-sight or a fixed plane to read tags as do traditional theft detection systems). The RFID gates at the library exit(s) can be as wide as four feet because the tags can be read at a distance of up to two feet by each of two parallel exit gate sensors.

Although employed in business and other industries, it was not much widely used in library. Recent studies [8] revealed that only 14% British libraries and 34 % of American libraries used DSS in library management decisions. Projects using DSS in library include early work by Bommer and Chorba. In 1984 Starlett Reidelbach and Hartse attempted to initiate it at university of Nebraska at Omaha .They reported that the attempt was partially successful. In 1991, Ottensmann and Gleeson developed a decision support system based on circulation of data to assist in public library acquisition budgeting and other library management decisions.

3 PROPOSED SYSTEM: COMBINATION OF RFID AND DSS

RFID and DSS technologies can prove beneficial to the library in their own way. Our System leverages the advantages of both RFID & DSS. We will go through the general working of the System in this section of proposed system. The figure below shows how the general working in the system will take place.



3.1 EXPLANATION

This system will enable the librarians to keep a record of all the books present in the library as well as provide an efficient DSS which will enable the librarian as well as the customers to make certain decisions. It will help the customers for faster access to the library facilities.

However, the customers need to carry RFID smart cards for logging in & out of the system. The books should have RFID tags embedded in them. Also, there will be a facility to send SMS to the customer to inform him/her of the returning date of the issued book.

It provides an easy solution to the librarian as well as the customers to keep track of library books and their respective stocks. This will help the user to search book easily according to their choice.

3.2 THE PROCESS

- Whenever a new book is acquired by the library, an RFID tag is to be attached into the book with the relevant information like, call number, accession number, book number, etc.
- The detailed information regarding the book is also captured in the computer database.
- The computer database also stores all information for individual users (patrons) of the library.
- Each patron will be supplied with a smart card. These smart cards carry identification data for each patron.

- When a patron needs to get a book issued, he can get it done without any manual intervention. As soon as the user will flash his smart card, system will automatically open his login account page.
- He can then place the selected books, one by one on the RFID scanner.
- The computer will automatically record all the data against his name.
- Similarly, when a patron wants to return books, he will simply place the books in the book drop counter near the scanner and the books automatically are adjusted for return against the patron's name. Thus, the library operations will be automated to a great extent. On the other hand, DSS will enable the librarians, as well as the patrons (users) to make decisions helpful to the library management .For instance, library owners can find the how frequently a particular book is issued .The frequently issued books can be ordered in large stocks compared to the other books. Thus, it can also prove beneficial in an economic way.

Data mining, which is a part of DSS, will provide the library users with a large range of choice. Say, if the user, who is logged into the system, enters the name (or a particular set of keywords) of the book of interest, the user will be provided with a list of some of the popular books based on the input entered. Then, he/she can select the book of his/her choice.

3.3 BENEFITS OVER EXISTING SYSTEM

The proposed system is expected to be more efficient compared to the existing system .As it is an integration of the two technologies , RFID and DSS ,their individual benefits will be combined here.

3.4 ADVANTAGES OF RFID

RFID application in libraries will benefit all persons involved in managing, running and using their facilities. One of the major benefits of an RFID system in a library is the ease of check-in and check-out of library items. Patrons can self check-in and check-out library items, saving themselves valuable time.

RFID technology inventory and scanning of items:

- Will take only 10% or less time as compared to conventional systems
- Misplaced books and other materials can be found easily – the reader can hone-in of misplaced or wrongly shelved items quickly. Besides these incredible technological advantages, the RFID system in a library will offer the following as well:
 - RFID tags are safe for magnetic media such as CDs, DVDs, etc.
 - Some RFID tags are rewritable. If, for example, a cataloguing error occurs, it can be rectified quickly.
 - Less manual handling of items hence better preservation.

- Staff has more time from routine chores and can therefore provide better service to patrons
- Tags last longer than barcodes as reading is contact-less
- It will remove manual book keeping of records
- Less time consumption as line of sight and manual interaction are not needed for RFID-tag reading.
- Manual intervention will be minimized
- Manual errors will be minimized
- It will provide the long lasting labels
- It will provide fast searching of books

3.5 ADVANTAGES OF DSS

- DSS for Self Management, which will help the librarian for managing the self according to the interest of the people.
- DSS for Search of the likely Books, this part of the application will help the Member of the library who will search for the book of his interest to get an available choice.
- Data Mining will help the librarian, for searching in the large database of the library which may contain large Database for book information and also Database for Customer Info.
- DSS in Stock Management, it will help for managing the available Stocks for librarians, for ordering the newly launched Books and out dated stock.
- Customers Timeline alert, this will inform both Librarian and Customer about the last date of returning the Book and its penalties if any.

4 CONCLUSION

The striking feature of this paper is the combination of RFID, DSS and data mining in library management. It represents an excellent and unique combination of hardware and software which can be used in library.

RFID is a technology that is sparking interest in the library community. RFID technology is recommended for library automation. It will not only speed up book borrowing, monitoring, books searching processes but also free staff to do more user-service tasks .Mainly, manual intervention will be minimized which in turn, will reduce errors.

DSS on the other hand, can help the librarians to take certain decisions, which can also prove helpful in an economic way. It can also help the library users to select the books of their choice or interest, shelf management. Most libraries exist to serve the information needs of the users. Thus understanding those needs is crucial to library's success. Data mining can help to examine the individual user's behaviour which can aid in understanding that individual .It can also help in examining the behaviour of a large group of users for regular patterns. This can allow the library personnel to have a better idea of the information

needs of the user base, and therefore better customize the library services to meet those needs.

Thus, RFID and DSS can benefit libraries in their own way. But, the proposed system attempts to combine their benefits which promise to increase efficiency, productivity and enhance user satisfaction.

We are looking forward to make this system online to make even more accessible and convenient to the patrons .Also we are trying to introduce sms service to patrons as a part of deadline alert .Provisions can also be provided regarding GPS tracking to ensure stopping theft totally. Research is also being made in the field of decision support system to make the system much more predictable.

REFERENCES

- [1]LIBRARY AUTOMATION USING RFID - www.rfid4u.com/.../Library%20Automation%20Using%20RFID.pdf.
- [2] Karen Coyle, "Management of RFID in Libraries" , Preprint version of article published in the *Journal of Academic Librarianship*, v. 31, n. 5, pp. 486-489
- [3]INTERNATIONAL JOURNAL OF APPLIED ENGINEERING RESEARCH, DINDIGUL Volume 1, No1, 2010.
- [4] D-Lib Magazine March 2003 Volume 9 Number 3
- [5] New Zealand Digital Library project. Available at <<http://nzdl.org>>.
- [6] RFID Technology in Anna University Library Management: A Study by V Rajasekar M Arul Dhanakar R Pandian R Malliga
- [7]Boss. R. W, *RFID Technology for libraries (Monograph)*, Library Technology Reports, November– December 2003.
- [8]Wanda Dole and Chang Cherry S. " Consortium use of OCLC Collection Analysis CD: The Sunny Experience " , *Library resources and technical services*.
- [9] Psion Teklogix handheld reader manual – www.psiointeklogix.com
- [10] Mercury 4 RFID reader manual – www.thingmagic.com
- [11]UHF RFID – Libraries taking the next step into the future – www.Adilam.com.au
- [12]<http://www.capterra.com/library-automation-software/spotlight/8980/Surpass/Surpass%20Software?id=67825005>
- [13]<http://www.capterra.com/library-automation-software/spotlight/8980/Surpass/Surpass%20Software?id=67825005>
- [14]LibBest - www.libbest.com
- [15] <http://www.naclin.org/RFID%20FOR%20LIBRARY.ppt>

[16] www.rfid4u.com/.../Library%20Automation%20Using%20RFID.pdf

[17] Hooker, J. N. (1988) "A qualitative approach to logic inference." *Decision Support Systems*. 4: 45-69.

[18] Department of Management Science and Information Systems, Graduate School of Business, CBA 5.202, The University of Texas at Austin, Austin, TX 78712, USA, {sulin, abw}@bongo.cc.utexas.edu

[19] diuf.unifr.ch/ds/courses/dss2002/pdf/DSS.pdf

[20] <http://www.mmu.ac.uk/h-ss/dic/research/irr/index.htm> : Decision-making in libraries: DECIMAL. An Integrative Research Report

A Reduced Flooding Algorithm and Comparative Study of Grid Fisheye State Routing Protocol for MANET

S. Nithya Rekha and Dr. C.Chandrasekar

Abstract— Mobile Ad-hoc Network (MANET) is the self organizing collection of mobile nodes. The communication in MANET is done via a wireless media. Ad hoc wireless networks have massive commercial and military potential because of their mobility support. Due to demanding real time multimedia applications, Quality of Services (QoS) support in such infrastructure less networks have become essential. QoS routing in mobile Ad-Hoc networks is challenging due to rapid change in network topology. In this paper, we focused to reduce flooding performance of the Fisheye State Routing (FSR) protocol in Grid using ns-2 network simulator under different performance metrics scenario in respect to number of Nodes and Speed. A Comparative study of FSR and GFSR is made in Speed with metrics .For example, the connection establishment is costly in terms of time and resource where the network is mostly affected by connection request flooding. The proposed approach presents a way to reduce flooding in MANETs. Flooding is dictated by the propagation of connection-request packets from the source to its neighborhood nodes. The proposed architecture embarks on the concept of sharing neighborhood information. The proposed approach focuses on exposing its neighborhood peer to another node that is referred to as its friend-node, which had requested/forwarded connection request. If there is a high probability for the friend node to communicate through the exposed routes, this could improve the efficacy of bandwidth utilization by reducing flooding, as the routes have been acquired, without any broadcasts. Friendship between nodes is quantized based on empirical computations and heuristic algorithms. The nodes store the neighborhood information in their cache that is periodically verified for consistency. Simulation results show the performance of this proposed method.

Index Terms— MANET, Flooding, Fisheye State Routing(FSR) protocol, Grid FSR , cache, connection-request, sharing, friend-node, NS2, Performance Metrics, Speed.

1 INTRODUCTION

The principal objective of a routing protocol is efficient discovery and establishment of a route between the source and the destination so that there can be a timely and efficient delivery of information between them. A Locating Service is used to locate the receiver inside the network. It dynamically maps the logical address of the receiver to its current location in the network. Once the receiver is located, routing and forwarding algorithms are used to route the information through the MANET. The routing is done using one-hop transmission service provided by the enabling technologies to construct an end-to-end (reliable) delivery services, from sender to one or more receivers. A number of features are expected to be supported by the routing protocols which include parameters like minimal control & processing overhead, loop freedom & prevention, efficient dynamic topology establishment and maintenance, scalability, support for unidirectional links, security & reliability and support for Quality of Service.[4],[5].

The rest of this paper is organized as follows: First of all, we make a brief survey on FSR in section 2, with my previous research work. In section 3, proposed work in grid FSR to reduce flooding. Section 4, presents the Results and Discussion. Section 5, presents the Simulation Results of performance evaluation of various Parameters and section 6 concludes the paper.

2. RELATED WORK

In my previous research work,[1] the investigations was on the behavior of the Proactive Routing Protocol Fisheye State Routing (FSR) in the Grid by analysis of various parameters. The Performance metrics that are used to evaluate routing protocols are Packet Delivery Ratio (PDR), Network Control Overhead, Normalized Overhead, Throughput and Average End to End Delay. Experimental results reveal that FSR is more efficient in Grid FSR in all QOS constraints. FSR can be used in all Resource critical environments. Grid Fisheye state routing (GFSR) consumes less bandwidth by restricting the propagation of routing control messages in paths formed by alternating gateways and neighbor heads, and allowing the gateways to selectively include routing table entries in their control messages. PDR and Throughput are 100% efficient in Simulation Results in NS2.

- S.Nithya Rekha is currently pursuing Ph.D. in Computer Science as a Full-Time Research Scholar in Periyar University, India, PH- +919944954001. E-mail: rekhasiva24@gmail.com
- Dr. C.Chandrasekar is an Associate Professor in Dept. of Computer Science in Periyar University, India, PH- +919994599967. E-mail:ccsekar@gmail.com

2.1 FSR (Fisheye State Routing)

G. P. Mario, M. Gerla and T.W. Chen, (2000) et.al, proposed that the FSR is a descendant of GSR [2]. In [3], the authors introduce a novel proactive (FSR), the notion of multi-level fisheye scope to reduce routing update overhead in large networks. Nodes exchange link state entries with their neighbors with a frequency which depends on distance to destination. From link state entries, nodes construct the topology map of the entire network and compute optimal routes. Simulation experiments show that FSR is simple, efficient and scalable routing solution in a mobile, ad hoc environment. Fig. 1 refers the fisheye scope with different hops.

The following are the advantages of FSR.

- * Simplicity
- * Usage of up-to-date shortest routes
- * Robustness to host mobility
- * Exchange Partial Routing Update with neighbors

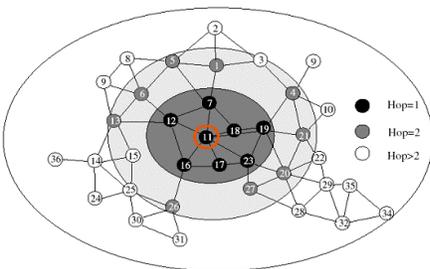


Fig. 1 : Fisheye Scope

3. PROPOSED WORK IN GRID FSR TO REDUCE FLOODING WITH RESPECT TO NODES AND SPEED AS PARAMETERS

The simulation results demonstrate the advantages of our approach. It can greatly reduce the redundant messages, thus saving much network bandwidth and energy. It can also enhance the reliability of broadcasting. It can be used in static or mobile wireless networks to implement scalable broadcast or multicast communications. As a result, the proactive approach provides a better quality of service by this new approach of Probability of calculating the Intimacy factor with neighbor node and friend node.

In Fig. 2, the gateways along the routing path will check the destination and determine who the next hop should be. When a packet approaches nearer to its destination, the routing information stored in the gateways becomes progressively more accurate. Packets finally arrive at the grid of destination node. If destination is a gateway, the packet is received by that gateway. Initially before data transmission each grid broadcast its grid member's information through gateway node. So that each gateway can exchange its grid members list as the whole network comes under the communication. While data transmission between grid members to other node gateway maintains unicast transmission until reaches its destination.

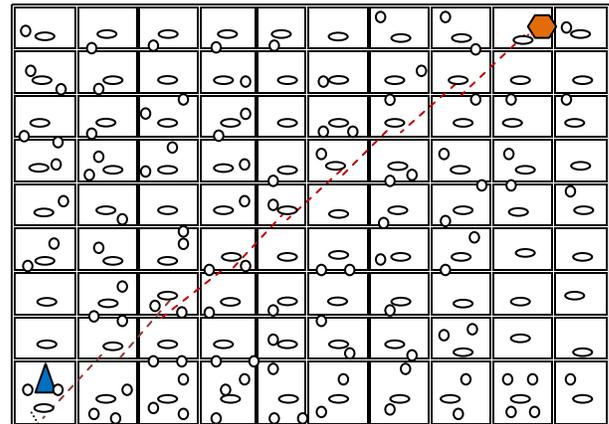


Fig. 2: GRID Architecture with Unicast data path from Source node to Destination Node

- Grid Gateway
- Grid Member
- ▲ Source Node
- ◻ GRID
- ◼ Destination Node
- ◻ GRID
- Unicast Datapath

3.1. Caching & Timeouts

In order for the destination node to know the location of the destination or the receiver they have to acquire the route through the process of flooding. Flooding involves broadcasting of a packet to all the nodes of the network requesting the route of the destination node [6]. The nodes either respond with a reply back to the sender if in case, the current receiving node of the packet is the destination, or otherwise they forward the packet to other nodes. The destination node responds to the sender with the connection acknowledgement. The path traced by the acknowledge packet is remembered in all the forwarding hosts as the route from sender to destination.

On one hand though the process of flooding helps the sender to dynamically obtain the location of the destination and the route over which information could be transmitted, it unnecessarily augments the load on the network as all the nodes in the network participate in the process of flooding[5]. If flooding is carried out frequently then it may altogether lead to the instability of the network. The direct implication of this observation is that flooding should be kept as infrequent as possible. One standard way to reduce the flooding mechanism is to provide the nodes with a small cache where the routes could be stored for future reference. The continuously changing characteristic of the ad hoc network environment poses further a problem, when routes are stored in a cache. There is always a possibility for the destination nodes to move from their place to another or even switch off. The cache reflects a static value and therefore in order to keep the data in the cache consistent they must be updated as frequently as possible. To ac-

count for these the routes in cache must be updated and validated periodically. The direct implication of this is that broadcast will be done frequently, which is to be avoided at all costs.

A new parameter referred to as the timeout period[6] is introduced to alleviate the problems arising. The timeout period is maintained for every route of the destination stored in the cache. This parameter reflects the lifetime of a route. The moment the timeout value expires the route in cache is deleted. Though still broadcasts are made, the frequency of broadcasts is reduced. The value of the timeout period reflects the frequency at which flooding occurs and if it is chosen to be a large number there is still a possibility of the route to have become invalid before the expiry of the cache. Therefore, the timeout value has to be prudently chosen. A tradeoff has therefore to be struck between the consistency of information and the frequency of broadcasts made.

3.2 Neighborhood

The suggested protocol falls under the category of reactive protocols. The nodes in this approach obtain the routes only when demand arises. The nodes use the common flooding approach to acquire the routes. Though flooding is used at the initial phases, it is decreased gradually. The most common method of using a cache with a fixed timeout period for each route is also used. The nodes are equipped with a small cache to save the routes and a timeout value is chosen. The selection of the timeout value is done as in case of the conventional networks, keeping in mind the other constraints of the network. However, the variation in the approach comes from the fact that the expiry of the timeout period does not trigger an update. The routes of the destination in the cache are rather erased after the timeout period. The nodes may then have to use flooding again to regain the routes, but in order to avoid that routes are shared between the nodes based on some criteria.

The primary focus of the protocol is on sharing information about the neighborhood of a peer with yet another node in the network. The neighborhood of a peer here reflects the contacts of the node in question with that of other nodes in the network. In other words, the neighborhood reflects the entries of routes in the cache. This information regarding the contacts a node has with other nodes in the vicinity is stored in tables or any other suitable data structure that is compatible with the protocol being adopted. The sharing of neighborhood information is not a mandatory task rather it is done at the discretion of the nodes concerned. The given approach intends to minimize the flooding requests that are needed to acquire the same information in the absence of the sharing mechanism.

The sharing of information occurs mainly between the nodes that are in direct contact with one another. Though the same can be carried out between nodes that are connected through a series of a finite number of intermediate nodes, factors such as the power levels of the nodes could be questioned to decide whether such a sharing could improve the efficiency or not. Since the sharing of information is at the discretion of the nodes in contact, they can decide whether the process has

to be carried out or not. The process affects only the nodes that are communicating and any series of finite set of intermediate nodes that connect the two nodes, if such a process is to be triggered between nodes connected by multi-hop links. The nodes can also take into account parameters such as the current load on the route connecting the nodes, the current load on the two nodes and the power levels required to carry out the operation before initiating the sharing of information.

3.3 Intimacy Factor

The process of sharing of neighborhood information occurs when the receiving node decides that the node that started the communication between the two is ready for accepting the information. This decision could be made based on a parameter called the intimacy factor. The intimacy factor reflects the level of trust between the two nodes that are communicating. A threshold level of intimacy factor could be defined called as IFTHRES, which could then be compared against the intimacy factor, calculated between two communicating nodes to determine, when exactly to commence the sharing of neighborhood information. If the intimacy factor calculated is greater than IFTHRES then the receiving node can make a request to the sender enquiring its acceptance to the information about the routes to nearby nodes. This request is optional and the receiver does so at its discretion.

After the receiver ensures that the node that initiated the communication is ready to receive the neighborhood information, it posts a request to the sender. The sender can accept or reject the request. It can take into account the load on the link, the load on it, and its power level before posting to the receiver its consent. This can ensure that the sharing of routes doesn't exhaust the limited resources available. After the transmission of the sender's consent to the request, the sharing of the information or routes begins. The receiver shares a percentage of its cache entries with its friend node, depending upon the power levels and other such criteria. The sender then comes to know of the locations good possibility for the sender to send messages to these of various destinations close to the receiver. There may be a destinations, in which case the flooding process required for acquiring the same, have been eliminated.

3.4 Designing Approach

In a MANET, the presented approach could be modeled in the following way.

$$\begin{aligned} \text{Total number of nodes in the network} &= T_n \\ \text{Total number of nodes in cache} &= K_n \\ \text{Unknown nodes} &= U_n \end{aligned}$$

The network is considered to have T_n number of nodes. The initiator of communication or the sender is assumed to have knowledge of routes of certain number of nodes in the network. The sender is unaware of the route of the other nodes, of which some may be near the receiver, with which the sender is currently communicating. The receiver is assumed to have a

similar knowledge of routes of various nodes in the network. The set of routes in the receiver's cache need not be disjoint in comparison with the contents of the cache in the source's node; although the greater, the dissimilarity in the contents of the cache would imply a greater efficiency in the working of the protocol.

$$\text{Route Gain Ratio (RGR)} = \frac{\text{contents of sender's cache}}{\text{contents of receiver's cache}}$$

$$\text{RGR} \propto \eta$$

Where η , is the efficiency of the protocol.

After the receiving the routes of the nodes in the neighborhood of the receiver, these are stored in the cache of the sender. The basic understanding is that, given that the sender has contacted the receiver, it has a good probability to communicate with the nodes nearby the receiver. Since the approach is reactive protocol oriented, new routes have to be acquired before the transmission of information to the other destinations. Calculating the probability that the sender communicates with any of the unknown nodes or nodes for which it does not have the location, a clear understanding of the working efficiency of the protocol can be obtained.

Number of nodes (given) : T_n

Probability that an unknown node is contacted by the sender : P_u

The approach will prove to be efficient only if the sender can utilize the information obtained from the receiver before it expires.

Time available for the sender for utilizing the routes : T_{out}

Assuming the average time spent per node as, Average time spent in communicating with a node: T_{avg}

Total number of calls possible before routes expire :

$$T_{out} / T_{avg} = T_c$$

Total number of unknown nodes: U_n (nodes whose route are unknown to sender)

Proby. that an unknown node is contacted : P_u

$$P_u = (U_n C T_{calls}) / (T_n C U_n)$$

When T_n is large, P_n tends to be very small. The efficiency of the protocol increases only when the unknown nodes contacted for a subset of the nodes in the neighborhood of the receiver. In other words, the maximum efficiency is gained only when the unknown node contacted is one of those exposed by the receiver to the sender during the sharing of neighborhood information.

Let number of nodes exposed = E_n

Proby. that a node exposed is contacted : P_e

$$P_e = (E_n C T_c) / (U_n C E_n)$$

Probability that the node contacted forms a subset of the nodes exposed: $P = P_c * P_e$

If the probability that the node contacted is from the set of nodes whose routes have been exposed by the receiver, then the protocols succeeds in eliminating the flooding requests which otherwise would have been required to contact the unknown nodes. Considering the MANET environment to consist of a large number of nodes n and the probability P_u being small, Poisson distribution could used to model the situation as following.

Total number of nodes = $n = T_n$

Probability that an exposed node is communicated = P

Let x be the number of exposed nodes contacted by the sender. Then, $P = \lambda$.

The set of routes that are exposed are only valid until the timeout period, after which they are deleted from the cache. The quantity of maximum concern here is the number of exposed nodes that are contacted.

Proby. that x nodes are contacted = $P(X = x)$

$$P(X = x) = (e^{-\lambda} \lambda^x) / x!$$

$$P(X = x) = (e^{-np} \lambda^x) / x!$$

$$P(X = x) = (e^{-n(P_c * P_e)} \lambda^x) / x!$$

$$\text{Where } P_c = (U_c C T_c) / (T_c C U_n) \quad P_e = (E_n C T_c) / (U_n C E_n)$$

Total exposed nodes contacted: $T_e = P * E_n$

The higher the value of T_e , the lesser the broadcasts required for getting the routes for the unknown nodes. The probability that no exposed node is contacted is given by

$P(X = 0)$, where

$$P(X = 0) = e^{-n(P_c * P_e)}$$

$P_c * P_e > 0$ and always a finite quantity,

$$P(X = 0) = e^{-n(P_c * P_e)} = e^{-\lambda} > 0$$

3.5 Increasing the Probability

The probability of contacting an exposed node is therefore never zero. To improve the probability and decrease further the flooding process that are carried out, the value of $P(X=x)$ must be closer to unity.[6,10] To increase the number of exposed nodes contacted there exists two possible approaches, one by improving the value of E_n and the other wherein P is increased. Boosting the value of E_n is not under the control of the designer. E_n signifies the number of exposed nodes and is directly dependent on the neighborhood of the receiver that exposes the routes of the nodes to the sender. The value of E_n depends on the topology of the network, the density of the network and the mobility of the nodes in the network.

Although E_n is strictly not under the control of the network designer, the value of E_n can be enhanced considerably by increasing the number of nodes exposed. In general, the receiver might then be expected to expose routes of the direct contacts it has, to the sender. In order to escalate further the probability of contacting an exposed node, it can augment the sample space of the nodes exposed. In other words, it can expose more nodes. This involves the receiver exposing nodes that are connected to it even through multi-hop links. The different nodes can be exposed one by one based on priorities assigned to them according to the distance of the exposed node from the receiver. The sender may wish to stop the transaction at any time in the middle by issuing a "I'm satiated" message. The receiver on receiving the message stops sending the routes.

The second method of increasing the probability P to improve the value of T_e proves to be more feasible. In order to amplify the value of P the number of nodes that can be contacted before the exposed routes become invalid, can be boosted. This implies that the timeout period should be increased. If timeout value is enhanced then it can have two impacts on the network. The first impact is one, which would lead to lesser number of flooding, due to less frequent updates and a higher value of probability of contacting an exposed node.

The second would promote a chance for the data or the routes to be corrupted between the timeout periods. As a consequence of this, a tradeoff has to be struck between consistency of data and the reduction of flooding requests.

3.6 Cases

The total number of messages that are transferred between the sender and receiver depends on the amount of information shared between them. It also depends on the number of intermediate nodes that are present between the sender and the receiver. However, considering the fact that the sharing of information only affects the sender, receiver and the finite number of intermediate nodes, if any is present; it can be concluded that the number of messages processed and transferred would be less than in case of flooding. This relies on the fact that the latter process involves all the nodes of the MANET environment. Therefore, even if the sharing process is a slightly prolonged one the process does not have any impact on the other nodes of the network, which still remain free for communication.

Also in the suggested approach, the flooding requests are minimized to a hop count of one. The flooding is initially limited to the immediate circle of nodes remaining in the coverage region of the node that is broadcasting the flooding request. Of these nodes, if any has the route to the destination in its cache it posts a message to the sender of the broadcast request, replying that the destination is its friend. At this point, the sender might establish the connection with the responding node, rather than the destination nodes itself. This intermediate node then routes the messages sent to it, to the intended destination.

In the worst case, the immediate circle of nodes in the coverage region of the source node might not have the route to the destination. Under such circumstances, the source node can rebroadcast the message with a hop count that can be found using an algorithm to increase the depth of flooding exponentially. When the flooding is done again with a greater hop count than used in the previous broadcast, the request is posted to more number of nodes. The process is repeated with a more number of nodes covered during each time.

The connection to the intended destination is then broken into different connections that involve nodes in direct contacts. Therefore, the source sends the message to the node with which it maintains the direct contact and which knows the route to the destination. This node then contacts with the destination or with the other node in the set of nodes connecting the source with the destination. The focus of the sender of the information is then on passing the information packet only to the node in direct contact. The task of routing the packet to the intended destination is then vested with the intermediate node that receives the packet. This responsibility then shifts from intermediate node to another one, if multiple nodes are present between the link connecting the sender and the receiver, as in case of any other multi-hop link. The last node meets with the responsibility, when it receives the packet and transmits it to the intended receiver through a direct link.

3.7 Friend & Stranger Nodes

In general, when two nodes start communicating with each other the sender or the initiator of the communication is moved to the stranger node state with respect to the receiver. As the communication proceeds, the intimacy factor is augmented based on some well-defined method. After the intimacy factor crosses the threshold value, the stranger node moves to the friend node state again with respect to the receiving node. This transformation between the states indicates that the receiver now is starting to trust the sender and share some information regarding the routes of nodes in its vicinity. The change of state triggers the sharing of routes, which is initiated by the receiver at the end of the ongoing transactions. The speed of this state change is a very important parameter in the design of the protocol. The faster the change, the earlier the sender or the initiator obtains the neighborhood information. This also has the consequence of a malicious node being able to quickly get the location of various destinations and launch an attack on the network. After the state change, the receiver is identified as being ready to receive the request for sharing the information regarding nearby nodes. The nodes that are acquired from the receiver are stored in the cache with a timeout period. Like any ordinary route that is stored in the cache after the expiry of the timeout period as per the norms of the protocol the routes are cleared. The method of shifting the state of a source node or the initiator of a communication, from stranger node to friend node could be based either on some empirical or heuristic algorithms.

Empirically this could be done by maintaining a track of the messages transmitted between the nodes concerned or calculating the time during which the communication persists. It should also be noted that when the time of communication is taken into account, the factor could affect the sharing process. In fact, it could bring down the efficiency of the protocol as the time to make use of the routes acquired is reduced. A balance therefore must be found between the two parameters. On the other hand, if the factor is based on the messages transmitted, a counter must be maintained by the receiver to count the packets received. In the aforementioned situation, the counter value could be directly used as the intimacy factor or could be weighted by any suitable constant to give the intimacy factor values.

Let the number of packets transmitted by the stranger node to receiver by P_t .

$P_t \propto k^*$ Intimacy Factor, where k is some constant.

There also remains a good chance for the routes exposed to be already known to the sender. Under such circumstance, if possible the sender tries to correct the information that is maintained in the cache of the receiver. The sender then posts a "Gratis Reply" to the receiver. This informs the receiver the route, which was declared corrupt, and the new route that has to replace the corrupted one. A comparison is therefore required at the sender's side when it's receiving the exposed nodes' routes to ensure that the routes are correct. If during the comparison process the sender or the friend node to the receiver, identi-

fies a route that is already known to it but is different from the one exposed by the receiver, it has to be able to discriminate between the right and the faulty route. The faulty one need not always be a wrong route, but can be an old route for which a newer version exists. A mechanism can be used to either accept a standard reference or to communicate a chosen reference across nodes whichever proves feasible.

4. RESULTS AND DISCUSSION

The Network simulator 2 has been used to analyze the parametric performance of Fisheye State Routing Protocol (FSR) in Grid .The metric based analysis is shown in figure 3 to figure 14. We simulate flooding protocols using Network Simulator 2 [15].

Moreover, performance of flooding protocols using Grid FSR has reduced flooding with respect to nodes. The nodes are increased from 50, 75,100,125,150. Thus, the expectation that the efficient flooding scheme has improved the Grid FSR performance with various parameters.

4.1. Performance Metrics with Nodes

- End-to-End Delay: A specific packet is transmitting from source to destination and calculates the difference between send times and received times. Delays due to route discovery, queuing, propagation and transfer time are included in the delay metric. Certainly Fig. 3, and Fig.4 shows decrease in delay as in Flooding is reduced in FSR within Grid scenario.
- Packet Deliver Ratio (PDR): The (PDR) is defined as the ratio between the amount of packets sent by the source and received by the destination. Fig.5, and Fig.6 explains that PDR is more than 80% efficient in Reduced Flooding than FSR. On all other nodes PDR is better for FSR due to it scope technique and thus reduced traffic overhead.
- Throughput: Throughput is the average rate of successful data packets received at destination. It is usually measured in bits per second (bit/s or bps), and sometimes in data packets per second. The result are shown in Fig.7, and Fig.8
- Jitter: Jitter is the variation of the packet arrival time. In jitter calculation the variation in the packet arrival time is expected to be low. The delays between the different packets need to be low for better performance in ad-hoc networks. It becomes a matter of concern if it is more than the threshold value, which is different for data. The result are shown in Fig. 9 and Fig.10.
- Energy: Based on realistic simulation models, these protocols shows significant energy-conserving potential. Energy is reduced in flooding becomes paramount in constraining battery dimensions and replenishment needs. Although, many routing protocols that minimize the energy consumed for multi-hop packet delivery have been designed, most of them

surprisingly rely on flooding. The results are shown in Fig.11 and Fig.12 as energy is reduced due to disseminating flooding.

- Control Overhead: The results shows that overhead is reduced in Fig.13 and Fig.14. Network Control Overhead (NCO) [1] is used to show the efficiency of the MANET's routing protocol scheme. It is defined, as the ratio of the number of control messages (the number of routing packets, Address Resolution Protocol (ARP), and control packets e.g., RTS, CTS and ACK) propagated by each node throughout the network and the number of the data packets received by the destinations. The reductions of network control overhead at higher data rate are very significant. This is because the same amount of routing and control messages are needed.

5. SIMULATION RESULTS

5.1 Experimental Results with Nodes in Reduced Flooding

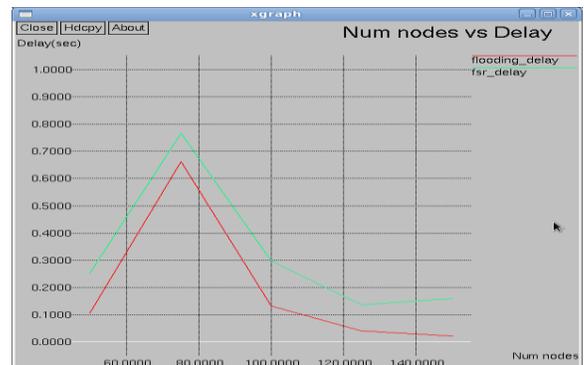


Fig. 3: End to End Delay V/s Nodes in Flooding

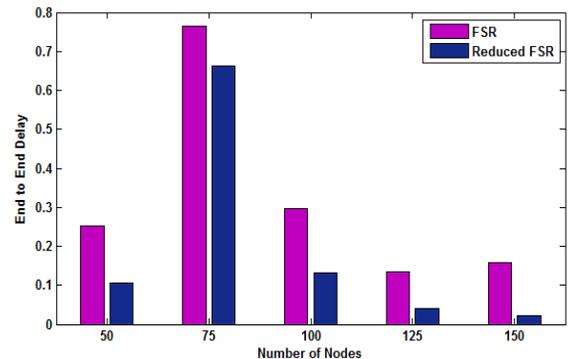


Fig. 4: Delay in Reduced Flooding with Grid FSR

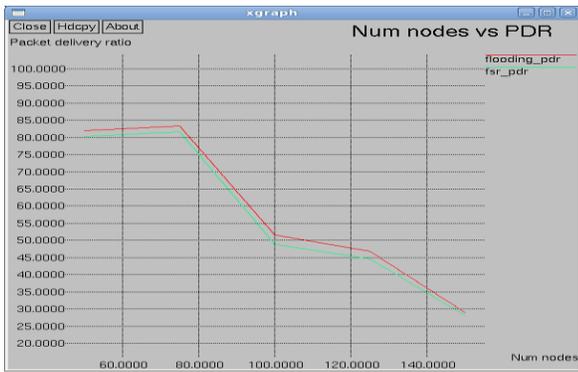


Fig. 5: Packet Delivery Ratio V/s Nodes in Flooding

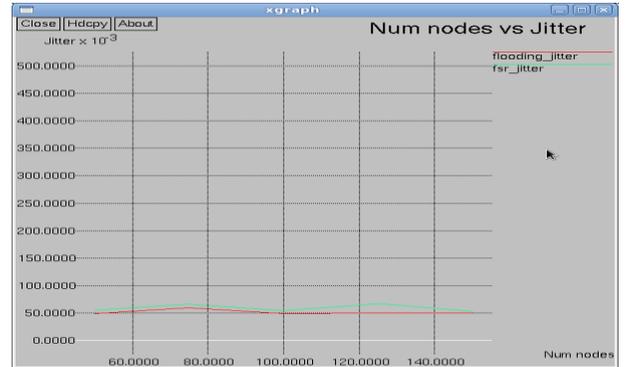


Fig. 9 : Jitter V/s Nodes in Reduced Flooding

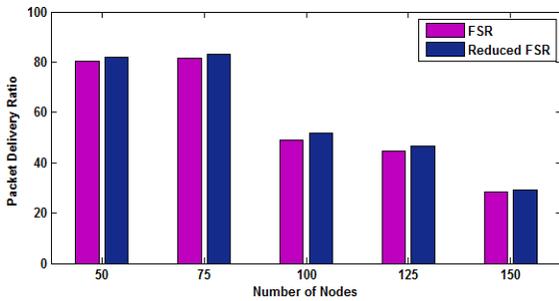


Fig. 6 : PDR in Reduced Flooding with Grid FSR

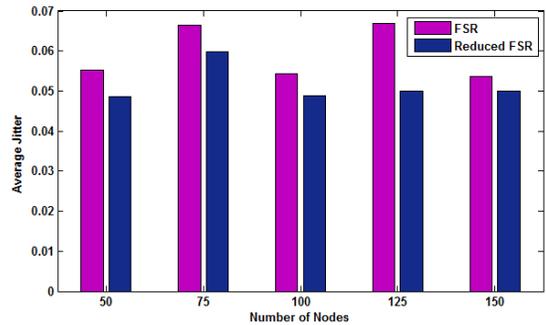


Fig. 10: Jitter in Reduced Flooding with Grid FSR

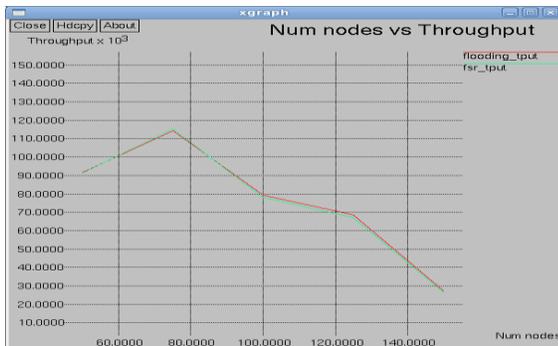


Fig.7: Throughput V/s Nodes in Reduced Flooding

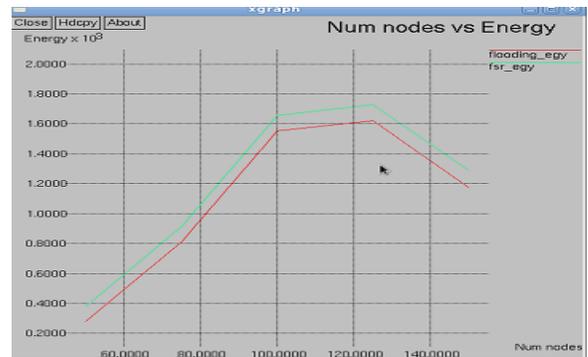


Fig. 11: Energy V/s Nodes in Reduced Flooding

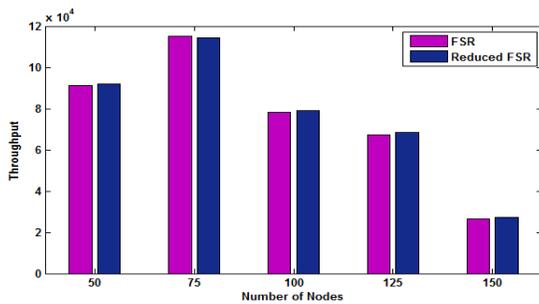


Fig. 8 : Throughput in Reduced Flooding with Grid FSR

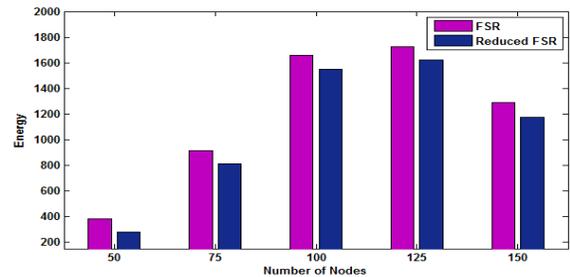


Fig. 12 : Energy in Reduced Flooding with Grid FSR

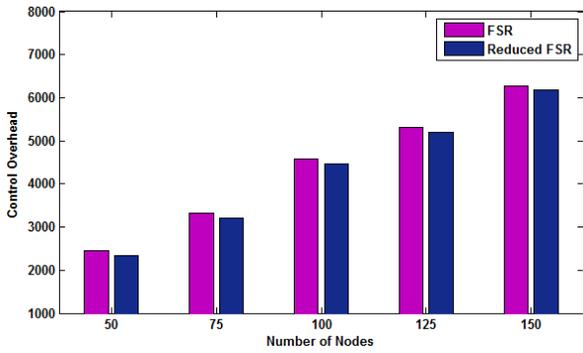


Fig. 13: Overhead in Reduced Flooding with GFSR

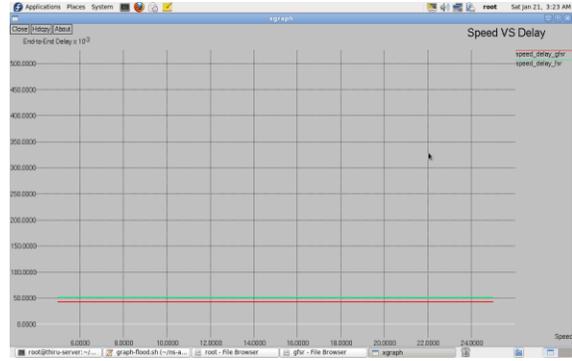


Fig. 16: FSR & GFSR – Speed V/s Delay

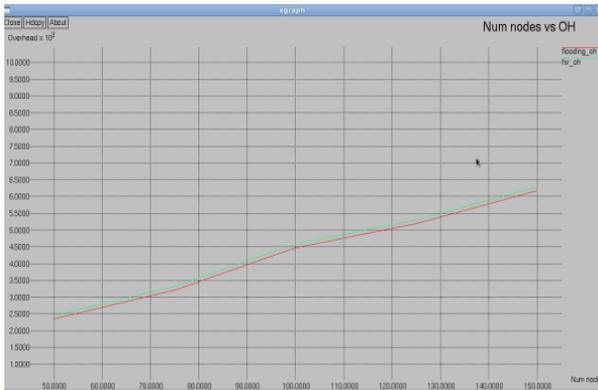


Fig. 14: Overhead V/s Nodes in Reduced Flooding

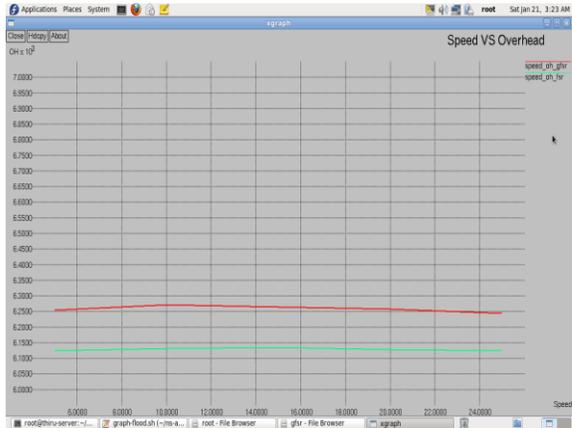


Fig. 17: FSR & GFSR – Speed V/s Overhead

5.2 Experimental Results with Performance Analysis and Comparison Study of FSR and GFSR with Speed in Reduced Flooding

In Fig. 15, GFSR have an increase in Throughput than FSR with respect to Speed. But there is a constant decrease in throughput with speed in FSR. Delay is reduced in GFSR flooding than FSR in Fig. 16. There is certain increase in GFSR Overhead than FSR in Fig. 17. The Speed is increased from 5, 10,15,20,25 m/s. in simulation.

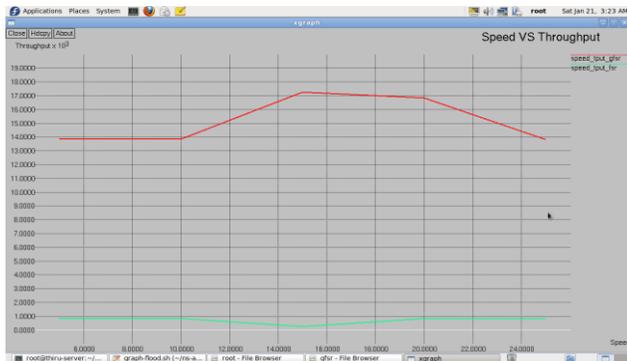


Fig. 15: FSR & GFSR – Speed V/s Throughput

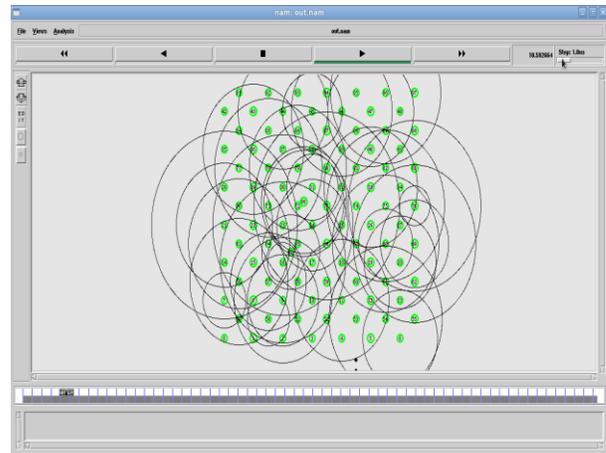


Fig. 18: Probabilistic Flooding of Rebroadcasting in Nam Window

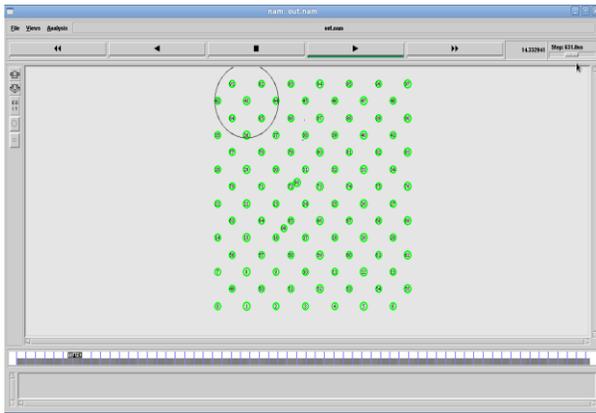


Fig. 19 : Reduced Flooding in NAM Window

6. CONCLUSION AND FUTURE WORK

Through a comparative study of efficient flooding, we have shown the above results. Figure 18 and 19 above shows the output NAM Window of Rebroadcasting and reduced flooding during simulation output.

Firstly, probability in Intimacy factor under certain circumstances, as the source node can rebroadcast the message with a hop count that can be found using an algorithm to increase the depth of flooding exponentially further to locate the destination. The protocol scales well for networks of all sizes. Secondly, The efficiency of the protocol has varying degrees of dependency with many parameters of the network, which includes the timeout period, the intimacy factor and others. Finally, the reduced flooding in routing traffic overhead and periodical propagation of link state information makes Grid FSR suitable for the high mobile dynamic changing network topology and thus the throughput is good with the high mobility of nodes, and therefore the average end-to-end delay is also very low. One of our future research works is to develop an efficient and optimized routing protocol with heavy mobility and routing overheads with different Mobility Models. Also the above proposed work will be interpreted in Rough Set Theory.

ACKNOWLEDGMENT

The First Author extends her gratitude to UGC as this research work was supported by Basic Scientist Research (BSR) Non-SAP Scheme, under grant reference number, F-4-1/2006(BSR)/11-142/2010(BSR) UGC XI Plan.

REFERENCE

[1] S.Nithya Rekha, Dr.C.Chandrasekar, R.Kaniezhil, "Efficient Routing Algorithm for MANET using Grid FSR", In the International Proceedings of Computer Science and Advancement in Information Technology & Business, Management and Governance (IPCSIT) 2011, Vol.20, pp.86-92, IACSIT Press, Singapore.

[2] T. W. Chen and M. Gerla, "Global State Routing: A New Routing Scheme for Ad-hoc Wireless Networks," In Proceedings of IEEE ICC'98, Atlanta, GA, Jun.1998, pp. 171-175.

[3] G. P. Mario, M. Gerla and T-W Chen, "Fisheye State Routing: A Routing Scheme for Ad Hoc Wireless Networks," Proceedings of the International Conference on Communications, New Orleans, USA, June 2000, pp. 70 -74.

[4] Ashish K. Maurya and Dinesh Singh, "Simulation based Performance Comparison of AODV, FSR and ZRP Routing Protocols in MANET", In International Journal of Computer Applications,(0975-8887) Vol.12-No.2, Nov.2010.

[5] Yunjung Yi, Mario Gerla, Taek Jin Kwon, Efficient Flooding in Ad hoc Networks: a Comparative Performance Study. In the Proceedings of the IEEE International Conference on Communications (ICC) 2003.

[6] I. Stojmenovic, (Handbook of Wireless Networks and Mobile Computing, Wiley, New York, 2002.

[7] S.Nithya Rekha, Dr.C.Chandrasekar, R.Kaniezhil, "Effect of Quality Parameters in Efficient Routing Protocol - Grid FSR with Best QoS Constraints" International Journal of Computer Applications, Vol. 37, No.2, pp.51-57, January 2012. Published by Foundation of Computer Science, New York, USA.

[8] Yoav Sasson, David Cavin, Andre Schiper, Probabilistic Broadcast for Flooding in Wireless Mobile Ad hoc Networks. In the Proceedings of Wireless Communications and Networking, 2003. WCNC 2003. IEEE Explore. Vol.2, pp. 1124 - 1130 .

[9] Jang-Ping Sheu, Cheng-Ta Hu, Chin-Min Chao, "Energy-Conserving Grid Routing Protocol In Mobile Ad Hoc Networks", Published In The handbook of ad hoc wireless networks by CRC Press, Inc. Boca Raton, FL, USA ©2003, ISBN:0-8493-1332-5.

[10] Ronald L. Graham, Donald E. Knuth, Oren Patashnik, Concrete Mathematics, Addison Wesley Publishing Company, Pub., 1989.

[11] M. Chatterjee, S.K. Das, and D. Turgut. WCA: A weighted clustering algorithm for mobile ad hoc networks. Journal of Cluster Computing, volume 5(2), pages 193-204, 2002.

[12] Y. Sasson, D. Cavin, and A. Schiper, Probabilistic broadcast for flooding in wireless mobile ad hoc networks, In Proc. IEEE Wireless Communications & Networking Conference (WCNC - 2003), pp. 1124-1130, March 2003.

[13] Ioannis C. Avramopoulos, Hisashi Kobayashi, "Guaranteed Delivery Flooding Protocols for Mobile Ad Hoc Networks".

[14] Imrich Chlamtac, Marco Conti, Jennifer J.-N. Liu, Mobile ad hoc networking: imperatives and challenges.

- [15] The Network Simulator ns-2, <http://www.isi.edu/nsnam/ns/ns-man.html>.
- [16] W. Lou, and J. Wu, On reducing broadcast redundancy in mobile ad hoc networks, *IEEE Transactions Mobile Computing*, Vol. 1, No. 2, pp. 111-123, 2002.
- [17] J. Wu and F. Dai, Broadcasting in ad hoc networks based on self-pruning, In *Proc. Annual Joint Conference IEEE Computer and Communications Societies (INFOCOM_2003)*, Vol. 3, pp. 2240-2250, 2003.
- [18] R. S. Sisodia, B. S. Manoj and C. Siva Ram Murthy, "A Preferred Link-Based Routing Protocol for Ad Hoc Wireless Networks," *Journal of Communications and Networks*, vol. 4, no. 1, March 2002, pp. 14-21.

Unsteady MHD flow of a Couple Stress Fluid Through a Porous Medium Between Parallel Plates Under the Influence of Pulsation of Pressure Gradient

M.Syamala Sarojini, M.Veera Krishna and C.Uma Shankar

Abstract— In this chapter, we discuss an analytical study of unsteady magneto hydro dynamic flow of an incompressible electrically conducting couple stress fluid through a porous medium between parallel plates, taking into account pulsation of the pressure gradient effect and under the influence of a uniform inclined magnetic field of strength H_0 inclined at an angle of inclination α with the normal to the boundaries. The solution of the problem is obtained with the help of perturbation technique. Analytical expression is given for the velocity field and the effects of the various governing parameters entering into the problem are discussed with the help of graphs. The shear stresses on the boundaries and the discharge between the plates are also obtained analytically and their behaviour computationally discussed with different variations in the governing parameters in detail.

Index Terms— couple stress fluids, inclined magnetic field, magneto hydro dynamic flows, porous medium, pulsation of the pressure gradient, parallel plate channels and unsteady flows

1 INTRODUCTION

A fluid flow driven by a pulsatile pressure gradient through porous media is of great interest in physiology and Biomedical Engineering. Such a study has application in the dialysis of blood through artificial kidneys or blood flow in the lung alveolar sheet. Ahmadi and Manvi [2] derived a general equation of motion for flow through porous medium and applied it to some fundamental flow problems. Rapits [8] has studied the flow of a polar fluid through a porous medium, taking angular velocity into account. The problem of peristaltic transport in a cylindrical tube through a porous medium has been investigated by El-Shehawey and El-Sebaei [7], their results show that the fluid phase means axial velocity increases with increasing the permeability parameter k . Afifi and Gad [1] have studied the flow of a Newtonian, incompressible fluid under the effect of transverse magnetic field through a porous medium between infinite parallel walls on which a sinusoidal traveling wave is imposed. The flow characteristics of a Casson fluid in a tube filled with a homogeneous porous medium was investigated by Dash et al [6]. Bhuyan Hazarika [4] has studied the pulsatile flow of blood in a porous channel in the presence of transverse magnetic field. The flows in bends and branches are of interest in a physiological context for several reasons. The additional energy losses due to the local disturbances of the flow are of interest in calculating the air flow in the lungs and in wave-propagation models of the arterial system.

The details of the pressure and shear stress distribution on the walls of a bend or bifurcation are of interest in the study of atherosclerosis because it appears that the localization of plaques is related to the local flow patterns. In vascular surgery questions arise, such as what is the best angle for vascular graft to enter an existing artery in a coronary bypass (Skalak, R. and Nihat Ozkaya, [12]). The theory of laminar, steady one-dimensional gravity flow of a non-Newtonian fluid along a solid plane surface for a fluid exhibiting slope at the wall has been studied by Astarita et al [3]. Suzuki and Tanaka [13] have carried out some experiments on non-Newtonian fluid along an inclined plane, the flow of Rivlin-Ericksen incompressible fluid through an inclined channel with two parallel flat walls under the influence of magnetic field has been studied by Rathod and Shrikanth [11]. Rathod and Shrikanth [9] have studied the MHD flow of Rivlin-Ericksen fluid between two infinite parallel inclined plates. The gravity flow of a fluid with couple stress along an inclined plane at an angle with horizontal has been studied by Chaturani and Upadhyaya [5]. Rathod and Thippeswamy [10] have studied the pulsatile flow of blood through a closed rectangular channel in the presence of microorganisms for gravity flow along an inclined channel. Hence, it appears that inclined plane is a useful device to study some properties of non-Newtonian fluids.

In this paper, we discuss an analytical study of unsteady magneto hydro dynamic flow of an incompressible electrically conducting couple stress fluid through a porous medium between parallel plates, taking into account pulsation of the pressure gradient effect and under the influence of a uniform inclined magnetic field of strength H_0 inclined at an angle of inclination α with the normal to the boundaries.

- Department of OR&SQC, Rayalaseema University, KURNOOL (A.P) - 518002 (INDIA) email: cumaor@rediffmail.com
- Department of Mathematics, Rayalaseema University, KURNOOL (A.P) - 518002 (INDIA) email: veerakrishna_maths@yahoo.com

2 FORMULATION AND SOLUTION OF THE PROBLEM

We consider the unsteady hydro magnetic flow of a couple stress fluid through a porous medium induced by the pulsation of the pressure gradient. The plates are assumed to be electrically insulated. The fluid is driven by a uniform pressure gradient parallel to the channel plates and the entire flow field is subjected to a uniform inclined magnetic field of strength H_0 inclined at an angle of inclination α with the normal to the boundaries in the transverse xy -plane.

We choose a Cartesian system $O(x, y, z)$ such that the boundary walls are at $z=0$ and $z=l$ and are assumed to be parallel to xy -plane. The equations for steady flow through porous medium are governed by Brinkman's model. At the interface the fluid satisfies the continuity condition of velocity and stress. The boundary plates are assumed to be parallel to xy -plane and the magnetic field of strength H_0 inclined at an angle of inclination α to the z -axis in the transverse xz -plane. This inclined magnetic field on the axial flow along the x -direction gives rise to the current density along y -direction in view of Ohm's law. Also the inclined magnetic field in the presence of current density exerts a Lorentz force with components along $O(x, z)$ direction, The component along z -direction induces a secondary flow in that direction while its x -components changes perturbation to the axial flow.

The steady hydro magnetic equations governing the couple stress fluid under the influence of a uniform inclined magnetic field of strength H_0 inclined at an angle of inclination α with reference to a frame are,

$$\frac{\partial u}{\partial t} = -\frac{1}{\rho} \frac{\partial p}{\partial x} + \frac{\mu}{\rho} \frac{\partial^2 u}{\partial z^2} - \frac{\eta}{\rho} \frac{\partial^4 u}{\partial z^4} - \frac{\sigma \mu_e^2 H_0^2 \text{Sin}^2 \alpha}{\rho} u - \frac{\mu}{k\rho} u \quad (1)$$

$$\frac{\partial w}{\partial t} = \frac{\mu}{\rho} \frac{\partial^2 u}{\partial z^2} - \frac{\eta}{\rho} \frac{\partial^4 u}{\partial z^4} - \frac{\sigma \mu_e^2 H_0^2 \text{Sin}^2 \alpha}{\rho} u - \frac{\mu}{k\rho} u \quad (2)$$

Where, the term $-\frac{\eta}{\rho} \frac{\partial^4 u}{\partial z^4}$ in the above equation gives the effect of couple stresses.

All the physical quantities in the above equation have their usual meaning. (u, w) are the velocity components along $O(x, z)$ directions respectively. ρ is the density of the fluid, μ_e is the magnetic permeability, ν is the coefficient of kinematic viscosity, k is the permeability of the medium, H_0 is the applied magnetic field.

Let $q = u + iw$

Now combining the equations (1) and (2), we obtain

$$\frac{\partial q}{\partial t} = -\frac{1}{\rho} \frac{\partial p}{\partial x} + \frac{\mu}{\rho} \frac{\partial^2 q}{\partial z^2} - \frac{\eta}{\rho} \frac{\partial^4 q}{\partial z^4} - \frac{\sigma \mu_e^2 H_0^2 \text{Sin}^2 \alpha}{\rho} q - \frac{\mu}{k\rho} q \quad (3)$$

The boundary conditions are, (Since the couple stresses vanish at both the plates which in turn) implies that

$$q = 0, \quad \text{at} \quad z = 0 \quad (4)$$

$$q = 0, \quad \text{at} \quad z = l \quad (5)$$

$$\frac{d^2 q}{dz^2} = 0, \quad \text{at} \quad z = 0 \quad (6)$$

$$\frac{d^2 q}{dz^2} = 0, \quad \text{at} \quad z = l \quad (7)$$

We introduce the non-dimensional variables

$$z^* = \frac{z}{l}, \quad q^* = \frac{ql}{\nu}, \quad P^* = \frac{Pl^2}{\rho\nu^2}, \quad t^* = \frac{t\nu}{l^2}, \quad \omega^* = \frac{\omega l^2}{\nu}, \quad x^* = \frac{x}{l}$$

Using the non-dimensional variables (dropping asterisks), we obtain

$$a^2 \frac{\partial q}{\partial t} + \frac{\partial^4 q}{\partial z^4} - a^2 \frac{\partial^2 q}{\partial z^2} + (M^2 \text{Sin}^2 \alpha + D^{-1}) a^2 q = -a^2 \frac{\partial p}{\partial x} \quad (8)$$

Where $a^2 = \frac{l^2 \mu}{\eta}$ is the couple stress parameter

$$M^2 = \frac{\sigma \mu_e^2 H_0^2 l^2}{\mu} \text{ is the Hartman number}$$

$$D^{-1} = \frac{l^2}{k} \text{ is the inverse Darcy parameter}$$

Corresponding the non-dimensional boundary conditions are given by

$$q = 0, \quad \text{at} \quad z = 0 \quad (9)$$

$$q = 0, \quad \text{at} \quad z = 1 \quad (10)$$

$$\frac{d^2 q}{dz^2} = 0, \quad \text{at} \quad z = 0 \quad (11)$$

$$\frac{d^2 q}{dz^2} = 0, \quad \text{at} \quad z = 1 \quad (12)$$

For the pulsation pressure gradient

$$-\frac{\partial p}{\partial x} = \left(\frac{\partial p}{\partial x} \right)_s + \left(\frac{\partial p}{\partial x} \right)_o \exp(i\omega t) \quad (13)$$

Equation (6.2.8) reduces to the form

$$a^2 \frac{\partial q}{\partial t} + \frac{\partial^4 q}{\partial z^4} - a^2 \frac{\partial^2 q}{\partial z^2} + (M^2 \text{Sin}^2 \alpha + D^{-1}) a^2 q = -a^2 \left\{ \left(\frac{\partial p}{\partial x} \right)_s + \left(\frac{\partial p}{\partial x} \right)_o e(i\omega t) \right\} \quad (14)$$

The equation (14) can be solved by using the following perturbation technique

$$u = u_s + u_o e(i\omega t) \quad (15)$$

Substituting the equation (15) in (14) and equating like terms on both sides

$$\frac{d^4 q_s}{dz^4} - a^2 \frac{d^2 q_s}{dz^2} + (M^2 \text{Sin}^2 \alpha + D^{-1}) a^2 q_s = -a^2 \left(\frac{\partial p}{\partial x} \right)_s \quad (16)$$

and

$$\frac{d^4 q_o}{dz^4} - a^2 \frac{d^2 q_o}{dz^2} + (M^2 \text{Sin}^2 \alpha + D^{-1} + i\omega) a^2 q_o = -a^2 \left(\frac{\partial p}{\partial x} \right)_o \quad (17)$$

Subjected to the boundary conditions

$$q_s = 0, \quad \text{at } z = 0 \tag{18}$$

$$q_s = 0, \quad \text{at } z = 1 \tag{19}$$

$$\frac{d^2 q_s}{dz^2} = 0, \quad \text{at } z = 0 \tag{20}$$

$$\frac{d^2 q_s}{dz^2} = 0, \quad \text{at } z = 1 \tag{21}$$

and

$$q_o = 0, \quad \text{at } z = 0 \tag{22}$$

$$q_o = 0, \quad \text{at } z = 1 \tag{23}$$

$$\frac{d^2 q_o}{dz^2} = 0, \quad \text{at } z = 0 \tag{24}$$

$$\frac{d^2 q_o}{dz^2} = 0, \quad \text{at } z = 1 \tag{25}$$

$$\left(\frac{\partial p}{\partial x}\right)_s = p_s \quad \text{and} \quad \left(\frac{\partial p}{\partial x}\right)_o = p_o$$

The solutions of the equations (16) and (17) subjected to the boundary conditions (18) to (25) give the velocity distribution of the fluid under consideration.

$$q = C_1 e^{m_1 z} + C_2 e^{m_2 z} + C_3 e^{-m_1 z} + C_4 e^{-m_2 z} + \frac{p_s + G \cos \phi}{M^2 \sin^2 \alpha + D^{-1}} + \left(C_5 e^{m_5 y} + C_6 e^{m_6 y} + C_7 e^{-m_5 y} + C_8 e^{-m_6 y} + \frac{p_o}{M^2 \sin^2 \alpha + D^{-1} + i\omega} \right) \exp(i\omega t) \tag{26}$$

Where, the constants C_1, C_2, \dots, C_8 are given in appendix.

The shear stresses on the lower and upper plates are given in dimension less form as

$$\tau_L = \left(\frac{dq}{dz}\right)_{z=0} = m_1(C_1 - C_3) + m_2(C_2 - C_4) + (m_5(C_5 - C_7) + m_6(C_6 - C_8)) \exp(i\omega t) \tag{27}$$

And

$$\tau_U = \left(\frac{dq}{dz}\right)_{z=1} = m_1(C_1 e^{m_1} - C_3 e^{-m_1}) + m_2(C_2 e^{m_2} - C_4 e^{-m_2}) + (m_5(C_5 e^{m_5} - C_7 e^{-m_5}) + m_6(C_6 e^{m_6} - C_8 e^{-m_6})) \exp(i\omega t) \tag{28}$$

The non-dimensional discharge between the plates per unit depth is given by Q

$$Q = \int_0^1 q(z,t) dz$$

$$= \frac{C_1}{m_1} (e^{m_1} - 1) + \frac{C_2}{m_2} (e^{m_2} - 1) - \frac{C_3}{m_1} (e^{-m_1} - 1) - \frac{C_4}{m_2} (e^{-m_1} - 1) + \frac{p_s}{M^2 \sin^2 \alpha + D^{-1}} + \left(\frac{C_5}{m_5} (e^{m_5} - 1) + \frac{C_6}{m_6} (e^{m_6} - 1) - \frac{C_7}{m_5} (e^{-m_5} - 1) - \frac{C_8}{m_6} (e^{-m_6} - 1) + \frac{p_o}{M^2 \sin^2 \alpha + D^{-1} + i\omega} \right) \exp(i\omega t) \tag{29}$$

3 RESULTS AND DISCUSSIONS

The unsteady state velocities representing the ultimate flow have been computed numerically for different sets of governing parameters namely viz. The Hartmann parameter M , the inverse Darcy parameter D^{-1} and couple stress parameter a and their profiles are plotted in figures (1-3) and (4-6) for the velocity components u and v respectively. For computational purpose we have assumed an angle of inclination α and the pulsation of pressure gradient in the x -direction and are fixed. Since the thermal buoyancy balances the pressure gradient in the absence of any other applied force in the direction, the flow takes place in planes parallel to the boundary plates. However the flow is three dimensional and all the perturbed variables have been obtained using boundary layer type equations, which reduce to two coupled differential equations for a complex velocity.

We notice that the magnitude of the velocity component u reduces and v increases with increasing the intensity of the magnetic field M the other parameters being fixed, it is interesting to note that the resultant velocity experiences retardation with increasing M (Fig. 1 & 4). (Fig. 2 & 5) exhibit both the velocity components u and v reduces with increasing the inverse Darcy parameter D^{-1} . Lower the permeability of the porous medium lesser the fluid speed in the entire fluid region. The resultant velocity experiences retardation with increasing the inverse Darcy parameter D^{-1} . Here we observe that the retardation due to an increase in the porous parameter is more rapid than that due to increase in the Hartmann number M . In other words, the resistance offered by the porosity of the medium is much more than the resistance due to the magnetic lines of force. We notice that u exhibits a great enhancement in contrast to v which retards appreciably with increase in the couple stress parameter S , but the resultant velocity shows and appreciable enhancement with in a (Fig. 3 & 6).

The shear stresses on the upper and lower plates and the discharge between the plates are calculated computationally and tabulated in the tables (1-5). The magnitude of these stresses at the upper plate is very high compared to the respective magnitudes at the lower plate. We notice that the magnitude of the both stresses τ_x and τ_y increase with increasing the couple stress parameter a on the upper plate and lower plates. On the upper plate, the magnitudes of τ_x and

τ_y increase with increasing M , but τ_x reduces and τ_y enhances with increase in D^{-1} , while on the lower plate τ_y rapidly enhances and τ_x reduces with increase in M . The reversal behavior shows that τ_x and τ_y with increase in D^{-1} (Tables. 1-4). The discharge Q reduces in general with increase in the intensity of the magnetic field M and lower permeability of the porous medium (corresponding to an increase in D^{-1}) and enhances the couple stress parameter a (Table. 5).

4 CONCLUSIONS

Under the effect of pulsation of pressure gradient, the resultant velocity experiences retardation with increasing M , where as the resultant velocity experiences retardation with increasing the inverse Darcy parameter D^{-1} . When we increase the couple stress fluid parameter, the resultant velocity shown and appreciable enhancement in the entire flow region. The magnitude of these stresses at the upper plate is very high compared to the respective magnitudes at the lower plate. The discharge Q reduces in general with increase in the intensity of the magnetic field M and lower permeability of the porous medium and enhances the couple stress parameter a

ACKNOWLEDGMENTS

We kindly acknowledge Prof. C. Uma Shankar and Dr. M.Veera Krishna, Department of Mathematics, Rayalaseema University, Kurnool (AP), India for their useful remarks on the manuscript and Prof. R. Sreenadh, Department of Mathematics, S.V.University, Tirupathi (AP), India for providing the material which was used to validate our computational work. Also, part of the computational facilities was provided by Department of Mathematics, S.K.University, Anantapur (AP), India.

REFERENCES

- [1] Afifi, N.A. S. and Gad (2001), "Interaction of peristaltic of peristaltic flow with pulsatile magneto-fluid through a porous medium". Acta Mechanica, vol. 149, pp. 229-237.
- [2] Ahmadi, G and Manvi, R (1971), "Equation of motion for viscous flow through a rigid porous medium", Ind. J. Tech. vol. 9, pp. 441-444.
- [3] Astarita, G., Mariucci, G. and Palumbo. G (1964), "Non-Bewtonian gravity flow on inclined surfaces". Ind. Eng. Chem. Fundam. vol. 3, pp. 333.
- [4] Bhuyan, B.C. and Hazarika, G.C.(2001), "Effects of magnetic field on pulsatile flow of blood in a porous channel", Bioscience Research Bulletin. vol. 17, Issue. No. 2, pp. 105-112.
- [5] Chaturani, P. and Upadhyya, V.S. (1977), "Gravity flow of a fluid with couple stress along an inclined plane with application to blood flow". Biorheology, vol.14, pp. 237-246.
- [6] Dash, R.K., Mehta, K.N. and Jayaraman.G (1996), "Casson fluid flow in a pipe filled with a homogeneous porous medium", Int. J. Eng. Sci. vol. 34. pp. 1145-1156.
- [7] El-Shehawey, E.F. and El-Sebaei, Wahed. (2000), "Peristaltic transport

TABLE 1
The shear stresses (τ_y) on the upper plate.

a ²	I	II	III	IV	V
0.25	1.085652	1.226565	1.566532	0.995682	0.457562
0.5	1.468656	1.566825	1.856825	1.246824	0.675645
0.75	1.533685	1.994526	2.085662	1.327564	0.766528
1	1.636899	2.246535	2.856824	1.457562	0.899465

	I	II	III	IV	V
M	2	5	8	2	2
D ⁻¹	1000	1000	1000	2000	3000

TABLE 2
The shear stresses (τ_y) on the upper plate.

a ²	I	II	III	IV	V
0.25	-0.45362	-0.63462	-0.85666	-0.66524	-1.56652
0.5	-0.76832	-0.79965	-0.99453	-0.94535	-2.00856
0.75	-0.99526	-1.00855	-1.45665	-1.47565	-2.24656
1	-1.28656	-0.38408	-1.88457	-1.83465	-2.56658

	I	II	III	IV	V
M	2	5	8	2	2
D ⁻¹	1000	1000	1000	2000	3000

TABLE 3
The shear stresses (τ_y) on the lower plate.

a ²	I	II	III	IV	V
0.25	0.028465	0.023265	0.018345	0.045652	0.065652
0.5	0.036652	0.033457	0.030085	0.095658	0.256656
0.75	0.042754	0.039942	0.032683	0.146524	0.846834
1	0.051156	0.045768	0.035683	0.284682	0.984653

	I	II	III	IV	V
M	2	5	8	2	2
D ⁻¹	1000	1000	1000	2000	3000

TABLE 4
The shear stresses (τ_y) on the lower plate.

a ²	I	II	III	IV	V
0.25	-0.03265	-0.03846	-0.04256	-0.01561	-0.00566
0.5	-0.04211	-0.04808	-0.05245	-0.02564	-0.01579
0.75	-0.05754	-0.06334	-0.06856	-0.04751	-0.02668
1	-0.06224	-0.06568	-0.07566	-0.05996	-0.03468

	I	II	III	IV	V
M	2	5	8	2	2
D ⁻¹	1000	1000	1000	2000	3000

Table 5
Discharge

a ²	I	II	III	IV	V
0.25	0.568898	0.502142	0.455789	0.487988	0.411254
0.5	0.685998	0.488755	0.388547	0.558264	0.501245
0.75	0.478871	0.322565	0.244587	0.445879	0.410025
1	0.356658	0.255478	0.189965	0.312455	0.322546

	I	II	III	IV	V
M	2	5	8	2	2
D ⁻¹	1000	1000	1000	2000	3000

in a cylindrical tube through a porous medium". Int. J. Math. & Math. Sci. vol. 24, Issue. no. 4, pp. 217-230.

[8] Rapits, A.(1982), "Effects of couple stress on the flow through a porous medium". Rheol. Acta, Vol. 21, PP. 736-737.
 [9] Rathod, V.P. and Hossurker Shrikanth, G. (1998), "MHD flow of Revin Ericksen fluid between two infinite parallel inclined plates". The Mathematics Education, XXXII (4), PP. 227-232.
 [10] Rathod, V.P. and Thippeswamy, G. (1999), "Gravity flow of pulsatile blood through closed rectangular inclined channel with micro-organisms". The Mathematics Education. XXXIII(1), pp. 40-49.
 [11] Rathod, V.P. and Hossurker Shrikanth, G. (1998), "MHD flow of Revin Ericksen fluid through an inclined channel". Bull. Of Pure & Appl. Sci., vol. 17 E(1), pp. 125-134.
 [12] Skalak, R. and Nihat Ozkaya, (2001), "Bio fluid Mechanics", Dept. of Civil Engin- eering and Engineering Mechanics, Coulumbia University, New York. p.167.
 [13] Suzuki, A. and Tanka, T. (1971), "Measurement of flow properties of powers along an inclined plane". I & EC Fundamental, vol. 10, p. 84.

Appendices

$$m_1 = \sqrt{\frac{a^2 + \sqrt{a^4 - 4a^2(M^2 \sin^2 \alpha + D^{-1})}}{2}}$$

$$m_2 = \sqrt{\frac{a^2 - \sqrt{a^4 - 4a^2(M^2 \sin^2 \alpha + D^{-1})}}{2}}$$

$$m_5 = \sqrt{\frac{a^2 + \sqrt{a^4 - 4a^2(M^2 \sin^2 \alpha + D^{-1} + i\omega)}}{2}}$$

$$m_6 = \sqrt{\frac{a^2 - \sqrt{a^4 - 4a^2(M^2 \sin^2 \alpha + D^{-1} + i\omega)}}{2}}$$

$$C_1 = - \left\{ C_2 + C_3 + C_4 + \frac{p_s}{M^2 \sin^2 \alpha + D^{-1}} \right\}$$

$$C_2 = \frac{-1}{e^{m_2} - e^{m_1}} \left\{ \left(e^{-m_1} - e^{m_1} \right) C_3 + \left(e^{-m_2} - e^{m_1} \right) C_4 + \left(\frac{p_s}{M^2 \sin^2 \alpha + D^{-1}} \right) \left(1 - e^{m_1} \right) \right\}$$

$$C_3 = -\frac{d_2}{d_1} C_4 - \frac{d_3}{d_1}, \quad C_4 = \frac{d_3 d_4 - d_1 d_6}{d_1 d_5 - d_2 d_4}$$

$$C_5 = - \left\{ C_6 + C_7 + C_8 + \frac{p_o}{M^2 \sin^2 \alpha + D^{-1} + i\omega} \right\}$$

$$C_6 = \frac{-1}{e^{m_6} - e^{m_5}} \left\{ \left(e^{-m_5} - e^{m_5} \right) C_7 + \left(e^{-m_6} - e^{m_5} \right) C_8 + \left(\frac{p_o}{M^2 \sin^2 \alpha + D^{-1} + i\omega} \right) \left(1 - e^{m_5} \right) \right\}$$

$$C_7 = -\frac{d_2'}{d_1} C_8 - \frac{d_3'}{d_1}, \quad C_8 = \frac{d_3' d_4' - d_1' d_6'}{d_1' d_5' - d_2' d_4'}$$

$$d_1 = \frac{(m_2^2 - m_1^2)(e^{m_1} - e^{-m_1})}{(e^{m_2} - e^{m_1})}$$

$$d_2 = (m_2^2 - m_1^2) \left\{ \frac{(e^{m_2} - e^{-m_2})}{(e^{m_2} - e^{m_1})} \right\}$$

$$d_3 = \frac{p_s}{M^2 \sin^2 \alpha + D^{-1}} \left\{ \frac{(m_1^2 - m_2^2)(1 - e^{m_1})}{(e^{m_2} - e^{m_1})} - m_1^2 \right\}$$

$$d_4 = \frac{e^{m_2} (m_1^2 - m_2^2)(e^{-m_1} - e^{m_1})}{(e^{m_2} - e^{m_1})}$$

$$d_5 = \frac{(m_2^2 e^{-m_2} - m_1^2 e^{m_1})(e^{m_2} - e^{m_1}) - (e^{-m_2} - e^{m_1})(m_2^2 e^{m_2} - m_1^2 e^{m_1})}{(e^{m_2} - e^{m_1})}$$

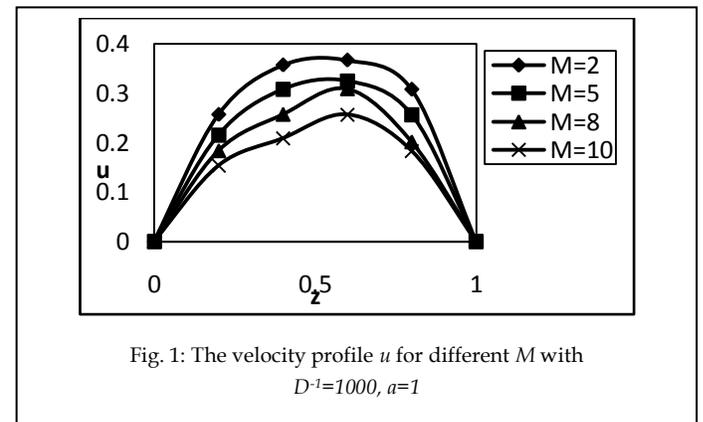
$$d_6 = \frac{p_s}{M^2 \sin^2 \alpha + D^{-1}} \left\{ \frac{(e^{m_2} - e^{m_1})(-m_1^2 e^{m_1}) - (1 - e^{m_1})(m_2^2 e^{m_2} - m_1^2 e^{m_1})}{(e^{m_2} - e^{m_1})} \right\}$$

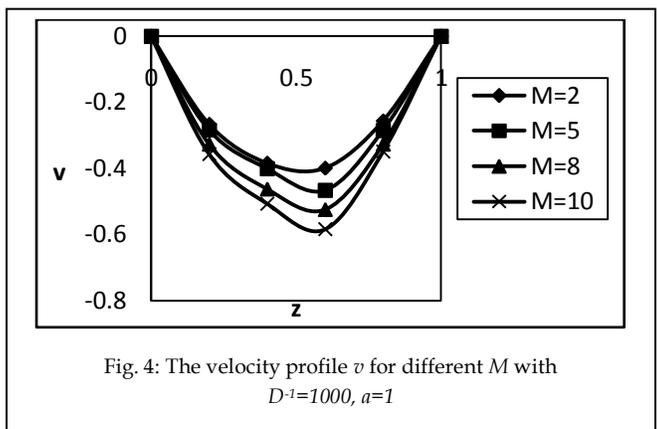
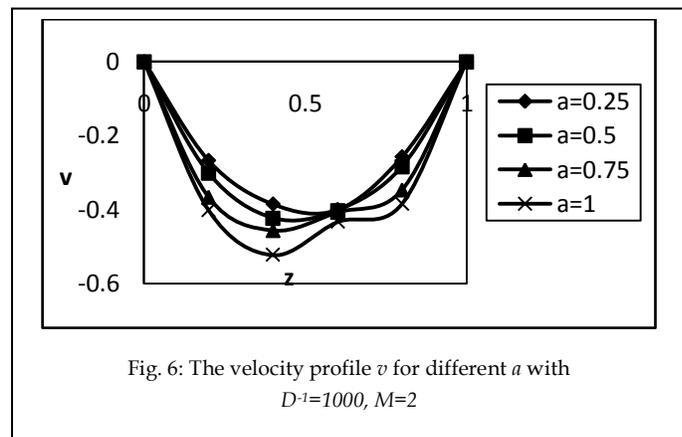
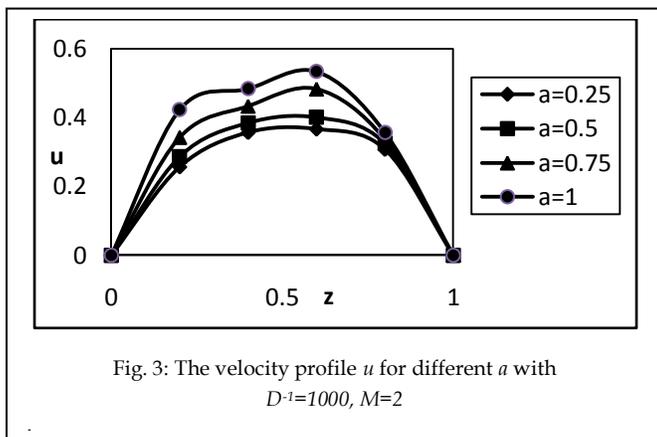
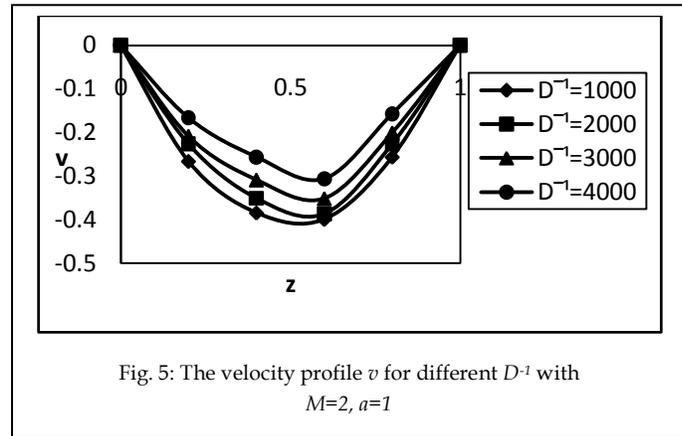
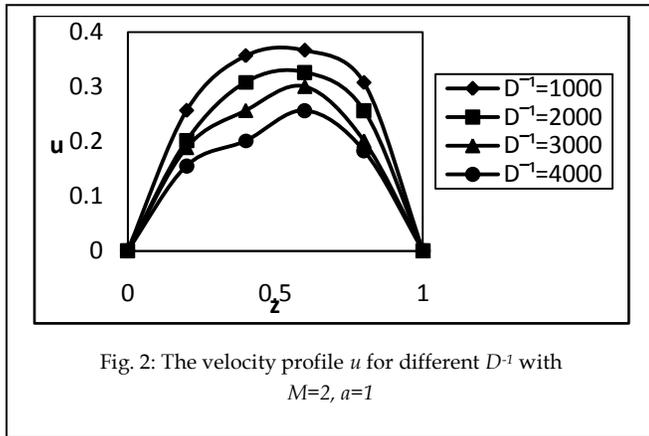
$$d_1' = \frac{(m_6^2 - m_5^2)(e^{m_5} - e^{-m_5})}{(e^{m_6} - e^{m_5})}, \quad d_2' = (m_6^2 - m_5^2) \left\{ \frac{(e^{m_6} - e^{-m_6})}{(e^{m_6} - e^{m_5})} \right\}$$

$$d_3' = \frac{p_o}{M^2 \sin^2 \alpha + D^{-1} + i\omega} \left\{ \frac{(m_5^2 - m_6^2)(1 - e^{m_5})}{(e^{m_6} - e^{m_5})} - m_5^2 \right\}$$

$$d_4' = \frac{e^{m_6} (m_5^2 - m_6^2)(e^{-m_5} - e^{m_5})}{(e^{m_6} - e^{m_5})}$$

GRAPHS





Innovative Launch Vehicle for Space Exploration

Achal Awasthi

Abstract— This paper presents various techniques to improve a space vehicle by using efficient means and procedures. The procedures aim at increasing the thrust, improving take off, design and landing, and using more efficient fuel in a space vehicle for better results.

Index Terms— boosters, fusion-powered, Lift-to-drag ratio, p-B11, polypropylene,

1 INTRODUCTION

Due to the scarcity of resources and lack of foolproof safety measures space flights are becoming more costly and rare. The Columbia disaster had shown that there is still scope for improvement in the field of space vehicles. These various propositions in propulsion, launching, design and landing not only help in making a space vehicle cost effective but also help in making it safe for astronauts.

2 PROPULSION

2.1 Existing technique

Spacecraft propulsion is any method used to accelerate spacecraft and artificial satellites. There are many different methods. Each method has drawbacks and advantages, and spacecraft propulsion is an active area of research. Currently spacecrafts are propelled by forcing a gas from the back/rear of the vehicle at very high speed through a supersonic de Laval nozzle. This sort of engine is called a rocket engine.

All current spacecraft use chemical rockets (bipropellant or solid-fuel) for launch, though some (such as the Pegasus rocket and Spaceship One) have used air-breathing engines on their first stage. Most satellites have simple reliable chemical thrusters (often monopropellant rockets) or resist jet rockets for orbital station-keeping and some use momentum wheels for attitude control. Interplanetary vehicles mostly use chemical rockets as well, although a few have used ion thrusters and Hall Effect thrusters (two different types of electric propulsion) to great success.

2.2 Proposed Technique

We can use fusion fuels which burn a mix of deuterium and tritium. The reaction produces most of its energy from neutron emissions, which tend to limit life of reactor and render the

whole structure radioactive. It also leads to environmental problems of burning fossil fuels. It is the fusion fuel you usually hear about because it is, the easiest reaction to produce, and it is the only fuel system with any chance at all of making net power from tokamaks, laser fusion.

We can use the p-B11 fusion fuel. It requires a temperature of nearly 6 billion degrees to burn in a system operating on Maxwellian heat which is vastly higher than for burning a mix of deuterium and tritium. This can be seen by taking a look at its reaction cross-section versus initiating energy graphs, but it is remarkably clean.

The Reaction:-

The reaction results in three alpha particles which when recombined with electrons again form Helium. The waste products of this reaction are not toxic and can be breathed without any significant side effects in humans and any creatures. The reaction produces almost no neutrons. Whereas natural boron is 80% B11, it is abundant and somewhat toxic. This reaction turns a toxin into an inert gas. It releases a large amount of energy. Alpha particles have a charge of +2. The first particle carries 43% of the reaction energy, and comes off (is ejected) at 3.76 MeV. The other two alphas come off (are ejected) at 2.46 million electron volts each. If you wanted to make an alpha with 3.76 MeV of energy, you will have to eject both electrons a helium atom, and accelerate it with an electric field of $3.76/2 = 1.88$ million volts. To get that energy back, simply decelerate that alpha particle against a 1.88 million volt field, let it gently touch a metal plate as it comes to a stop, and it will produce two electrons of current at that voltage. This has been done on a small scale using radioisotopes. All of the energy from this reaction comes off as alphas, and since their energies are relatively close together, it should be possible to devise a method of doing the same thing with the products of the p-B11 reaction. Setting the decelerating potential at $2.46/2 = 1.23$ million volts, would presumably recover 85% of the energy. Any nuclear reactor that generates its power as heat will wind up running steam turbines that waste 2/3 of the energy, this tech-

nology is quite useful. The environmental benefit of avoiding all that waste heat, the economic benefits of avoiding large cooling towers and implications for lightweight space propulsion systems all make this efficiency highly desirable. It is the greenest technology after photosynthesis. The big question?

Can electrodynamic fusion reactor burn boron? The answer surprisingly is yes. Boron has five electrons. Remove them all and the nucleus has a charge of +5. That means an electrostatic or electrodynamic acceleration system will work 5 times as hard on that nucleus as it would on a proton of charge +1.

The net result is that one only needs a potential well depth something like 100-150 kilovolts. So we can try to build a larger machine and run it at higher voltage. The world needs a technology that is compatible with existing power grids, affordable, compact, non-polluting, incapable of making nuclear weapons, and able to be used world-wide. If p-B11 fusion can be made to work; one can expect to see fusion-powered spacecraft.

3 LAUNCH

3.1 Existing technique

The space shuttle is launched in a vertical position, with thrust provided by two solid rocket boosters, called the first stage, and three space shuttle main engines, called the second stage. At liftoff, both the boosters and the main engines are operating. The three main engines together provide almost 1.2 million pounds of thrust and the two solid rocket boosters provide a total of 6,600,000 pounds of thrust. The total thrust at launch is about 7.8 million pounds. To achieve orbit, the shuttle must accelerate from zero to a speed of almost 28,968 kilometres per hour (18,000 miles per hour), a speed nine times as fast as the average rifle bullet. To travel that fast, it must reach an altitude above most of Earth's atmosphere so that friction with the air will not slow it down or overheat it. The journey starts relatively slowly: at liftoff, the shuttle weighs more than 2.04 million kilograms (4.5 million pounds) and it takes eight seconds for the engines and boosters to accelerate the ship to 161 kilometres per hour (100 mph.) But by the time the first minute has passed, the shuttle is travelling more than 1,609 kilometres per hour (1,000 mph) and it has already consumed more than one and a half million pounds of fuel by then.

3.2 Proposed Technique

The two boosters can be replaced by 4 boosters, two on either side since we have already discussed above a clean fuel that can be used which can provide a thrust of 13200000 pounds and will reduce the time needed for shuttle to accelerate. We can use a particle accelerator like a cyclotron which can accelerate the ejected electrons from the helium atom and thus can help in producing large amounts of energy which provide greater acceleration.

3.3 First Stage Ascent

3.3 a Existing Technique

After about two minutes, when the shuttle is about 45 kilometres (28 miles) high and travelling more than 4,828 kilometres per hour (3,000 mph), the propellant in the two boosters is exhausted and the booster casings are jettisoned. They parachute into some water body. These empty boosters are recovered to be eventually refilled with fuel and launched again. The solid fuel used by the boosters is actually powdered aluminium with oxygen provided by a chemical called ammonium perchlorate.

3.3 b Proposed Technique

1. The External Tanks are carried to orbit on each shuttle flight, then destroyed when force back down in the atmosphere when its odourless fuel is gone. We can keep them attached all the way into orbit and use these tanks for space station building blocks.
2. The boosters can be made up of high sunlight absorbing glass which can reduce the amount of cost required for fuels in providing thrusts.
3. This will also ensure that the boosters are not jettisoned after 2.26 minutes of the launch since they can absorb energy even after that and still help in providing thrust.

3.4 Second Stage Ascent

3.4 a Existing Technique

The three space shuttle main engines, attached to the rear of the shuttle orbiter, continue to fire until about 8.5 minutes after liftoff, burning a half-million gallons of liquid propellant from the large, orange external fuel tank as the shuttle accelerates. The main engines burn liquid hydrogen — the second coldest liquid on Earth at minus 252.7 degrees Celsius (minus 423 degrees Fahrenheit) — and liquid oxygen. Since the hydrogen and oxygen can reach a temperature as high as 3,315.6 degrees Celsius (6,000 degrees Fahrenheit) as they burn — higher than the boiling point of iron — the engines operate at greater temperature extremes than any other piece of machinery ever built. The engines' exhaust is primarily water vapour as the hydrogen and oxygen combine. Their turbines spin almost 13 times as fast as an automobile engine spins when it is running at highway speed. Eight and a half minutes after launch, with the shuttle travelling 8 kilometres (5 miles) a second, the engines shut down as they use the last of their fuel. A few seconds after the engines stop, the external fuel tank is jettisoned from the shuttle. The only part of the shuttle

that is not reused, the tank re-enters the atmosphere and burns up over the Pacific Ocean. The shuttle orbiter, the only space shuttle component that will circle the Earth, weighs only about 117,934 kilograms (260,000 pounds). The shuttle has consumed more than 1.59 million kilograms (3.5 million pounds) of fuel during its first 8 ½ minutes of flight.

3.4 b Proposed Technique

1. A cooling and electrolysis system can be installed in the main engine which can convert the water vapour obtained as the by-product by the burning of fuel into oxygen and hydrogen which can again be used as fuel in the main engine.
2. The above system should be made such that it can be operated manually, so when the astronauts feel they need an extra thrust or running out of fuel they can switch on this system. This way less amount of fuel will be used as the same fuel can be reused again and again.
3. By using this system the space vehicle will no longer be using orbits thus reducing its weight considerably. Since the boosters and main engines now can provide enough fuel to circularize the shuttle's orbit at a safe altitude to keep it above the atmosphere. (After the main engines shut down, the shuttle is in an egg-shaped orbit that, if nothing changed, would cause it to re-enter the earth's atmosphere. But, about 35 minutes after the main engines have shut down, usually when the shuttle has reached the highest point of the egg-shaped orbit, the two orbital manoeuvring system engines, located on the left and right side of the shuttle's tail, are fired for about three minutes. The orbital manoeuvring system engines use two propellants that automatically burn whenever they contact one another and the three-minute firing circularizes the shuttle's orbit at a safe altitude).

Reducing the cost

Most of the launch cost is due to weight and the fuel required for launching the vehicle to the proper orbit. If additional thrust could be created once in space, less fuel would be required, thereby also reducing weight factors.

What new can be done?

The boosters can be made up of glasses which can absorb large amounts of sunlight inside which an energy absorbing system can be kept. The absorber can then be cooled by heating hydrogen gas circulating through it. The hydrogen which in turn becomes hot can be expanded through a nozzle to provide very high thrust.

4 DESIGN

Various designs have been proposed about space vehicles.

What I propose is a design where:

1. The Lift-to-drag ratio or L/D close to 7 or slightly less. A lift to drag ratio of 7 means that a thrust force equal to 1/7th of the weight of the aircraft is sufficient to support it in flight. This low thrust requirement significantly reduces the amount of fuel required to carry the weight of an aerospace plane in comparison to rocket launch systems which must provide thrust greater than the weight of the vehicle.
2. A wingless launch vehicle has lower aerodynamic forces affecting the vehicle. Some fins are attached to aid stability. For a winged vehicle the centre of lift moves during the atmospheric flight as well as the centre of mass and the vehicle spends longer in the atmosphere as well. Thus a modern space vehicle should have a wide wingspan to provide extra stability and the wings should have minimal thickness. The wings should be as light as possible.
3. The engines should preferably be at the front end rather than the rear one and should be as light in weight as possible as having a heavy engine at the rear end puts a heavy mass at the rear of the aircraft with wings that had to hold up the vehicle. As the wet mass reduces, the centre of mass tends to move rearward behind the centre of lift, which tends to be around the centre of the wings. This can cause severe instability which is usually solved by wings but they also add weight and decrease performance.
4. A vertically-launched rocket forms the shape of a cylinder stood on end. This structure can be made very light and strong. A horizontally-launched space vehicle approximates a cylinder on its side. This structure experiences greater bending forces, so must be strengthened. This makes it heavier.
5. To reduce its weight one can attach the fuel tank directly to the space vehicle thus saving the weight of fasteners while also stiffening both parts.
6. One can create a space vehicle with a material like polypropylene which is currently used to cover ice sheets in Greenland and can withstand very high and low temperature.

5 LANDING

5.1 Existing technique

When it is time to return to Earth, the shuttle is rotated tailfirst into the direction of travel to prepare for another firing of the orbital maneuvering system engines, a firing called the deorbit burn. This three minutes long engine firing slows the shuttle and it begins to descend toward the atmosphere. The three-minute firing is the only active brake the shuttle will use as it heads toward a landing. The rest of its descent is devoted to slowing down using only the drag produced by the atmosphere. After the firing takes place, it is about another 25 minutes before the shuttle will descend to a point that it first en-

counters the effects of the atmosphere. Before it reaches the upper atmosphere, the shuttle is oriented with the nose angled up about 40 degrees from horizontal and its wings level, an orientation that keeps the black thermal tiles on the underside facing the majority of the heat generated by its encounter, heat that can range as high as 1,648.9 degrees Celsius the leading edges of the wings and nose. The aft steering jets are used to control the shuttle's orientation as it descends into the atmosphere. As it descends, however, it begins a transition from spacecraft to aircraft, and its aero surfaces -- the wing flaps and rudder -- gradually become active as air pressure builds. As those surfaces become usable, the steering jets turn off automatically. During its descent, the shuttle performs a series of four steep banks, rolling over as much as 80 degrees to one side or the other, to slow down. The series of banks gives the shuttle's ground track toward landing an appearance similar to a highly elongated letter "S".

As the shuttle continues toward landing and its speed drops to less than three times the speed of sound, or Mach 3, two air data probes are deployed from either side of the nose of the spacecraft. These probes provide supplemental information on the airspeed and altitude derived from the outside barometric pressure and wind speed.

As it aligns with the runway, the shuttle then begins a steep descent with the nose angled as much as 19 degrees down from horizontal. At this point; the pilot deploys the landing gear. As the shuttle's main landing gear touches down, it is dropping at less than 8 kilometres per hour (5 miles per hour) and has a forward speed of about 354 kilometres per hour (220 miles per hour). After touchdown, the pilot deploys a drag chute from a compartment located just below the tail and the commander begins to drop the shuttle's nose gear slowly toward the runway. The drag chute is then jettisoned before the wheels come to a stop to ensure that it falls clear of the shuttle.

5.2 Proposed Technique

The fuel present in the space vehicle can be reignited and made to burn at times, during descend. When the vehicle enters the earth's atmosphere it should be made to rotate several times and the burning of fuel should be used to align itself quickly. This will reduce the time taken for its landing.

Also the vehicle can be coated with a coolant so that when it is re-entering earth's atmosphere it does not catch fire due to overheating. These are some of the new steps that can be taken to ensure safe and fast landing of space vehicles.

6 CONCLUSION

This study shows that more efficient and non polluting fuels can be used to reduce the cost of a space vehicle. Its ascent, descent and landing could be made safer by trying to implement one the the aforementioned techniques. This could make traveling to space safer for astronauts and more cost effective for organizations

7 REFERENCES

WIKIPEDIA

Wavelike Properties of Time

Goutam Kumar Chandra

Abstract —Time is a component of the measuring system used to sequence events, to compare the durations of events and the intervals between them. The measuring equipment is likely to be a clock. By making an assumption that clock is measuring Time, it is observed experimentally that “Time exhibit wavelike properties “.

Index Terms— Special Relativity , Philosophy of Science , Time and Frequency

1. INTRODUCTION

Properties of Time can only be identified if we measure relative Time between two clocks. An experiment has been carried out by synchronizing two second accurate clocks at the same time. Both the clocks traversed different space-time compared to each other over the duration of one month and Δt being measured repeatedly within some finite time intervals.

2. EXPERIMENT

Let's represent two clocks as A and B. Initially B was kept 3 seconds ahead of A, so $\Delta t=3$. B used to traversed more space-time compared to A, but the space-time difference between B and A was minute in large-scale. Δt was found to be a nondeterministic wave function over the period of time.

3. MATHEMATICAL FRAMEWORK

$f_1(t)$ and $f_2(t)$ be the time function of B and A.

$$f_1(t) - f_2(t) = \Delta t \quad (1)$$

Δt is a time dependent wave function, shown in Fig.1.

Lemma: If Δt is a time dependent wave function, $f_1(t)$ and $f_2(t)$ also be a wave function.

4. CONCLUSION

Can we conclude Time exhibit wave properties?

G.K. Chandra is an Engineer based in India,
E- mail: goutam.k.chandra@gmail.com

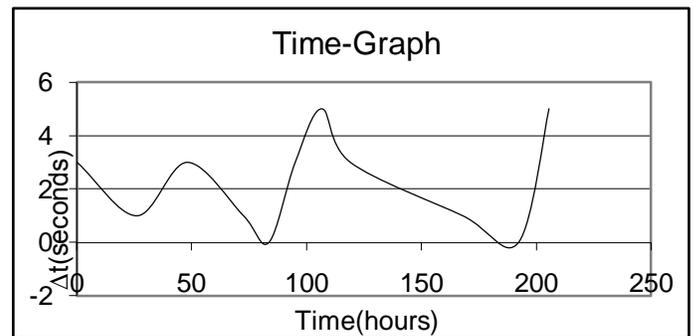


Figure.1 Time(in hours) is plotted along the X axis and Δt (in seconds) is plotted along the Y axis.

Dedication:

To the late memories of my Father, Mr. Aswini Kumar Chandra(1934-2006)

References:

[1] Wikipedia

INTRODUCING P2P IN CLOUD COMPUTING

B.S Sarada

Assistant Professor

Department of Computer Science

Mar Baselios College of Engineering & Technology

E-mail ID: sarada_bs@yahoo.co.uk

Abstract

Cloud computing platform is a set of scalable large-scale data server clusters, it provides computing and storage services to customers. The basic study about the architecture of current cloud computing system shows that it's a central structured one; i.e. All the data nodes are indexed by a master server, but when the number requests increases it may become bottle neck of the system. This research paper is about a cloud storage architecture based on P2P with fault tolerance. Here the central entity is removed and all servers are interconnected to form a ring structure. When one of the servers fails, the work will be taken over by any of the best performing servers. The confidentiality and integrity of data passed in between the servers is also maintained using MAC algorithm.

Keywords— cloud computing, fault tolerance, P2P, storage.

I. INTRODUCTION

Cloud computing is referred to as fifth generation computers. It sounds like a great working environment for non-technical people, as it takes away the burden of installing software's, increasing the memory capacity. It is a pay for use system, as it helps the users to access any servers from the pervasive network. It takes way the burden of managing huge networks. Cloud computing is a computing paradigm shift where computing is moved away from personal computers or an individual server to a "cloud" of computers. This method of distributed computing is done through pooling all computer resources together and being managed by software rather than a human. Cloud

computing lets you access all your applications and documents from anywhere in the world, freeing you from the confines of the desktop and making it easier for group members in different locations to collaborate. With cloud computing, the software programs that are used aren't run from own personal computer, but are rather stored on servers accessed via the Internet.

II. LITERATURE SURVEY

Cloud computing is a term used to describe both a platform and type of application [3]. A cloud computing platform dynamically provisions, configures, reconfigures, and deprovisions servers as needed. Servers in the cloud can be physical machines or virtual machines. Advanced clouds typically include other computing resources such as storage area networks (SANs), network equipment, firewall and other security devices. Cloud computing also describes applications that are extended to be accessible through the Internet. These cloud applications use large data centers and powerful servers that host web applications and web services. Anyone with a suitable Internet connection and a standard browser can access a cloud application.

The key features of the cloud computing are:

- The cloud computing design is done in such a way that the resources present in it will be available from anywhere at anytime.
- Since replication of data is done in cloud computing, the resources are available even during hardware failure.
- Cloud computing provides greater speed in its operation.
- The on- demand application deployment increases the resource utilization to a large extend.

- Low cost servers are available for storage and services.

Cloud computing system [2] has the capability to hold heavy load situations without much hardware support. It makes use of the virtualization concept. For client based transactions its better to store the data in cloud. By storing the data in the cloud the traditional productivity nature is preserved and also the new upcoming technologies could also be integrated. The cloud computing architecture fosters the innovative nature, then by encouraging the new technologies to develop. The cloud architecture is highly beneficial for both large scale and small scale industries. The large scale business firms also started moving to cloud due to the features like ease of availability, remote access, cost reduction.

Google File System [4] is a scalable distributed file system for large distributed data-intensive applications. It provides fault tolerance while running on inexpensive commodity hardware, and it delivers high aggregate performance to a large number of clients. While sharing many of the same goals as previous distributed file systems, our design has been driven by observations of our application workloads and technological environment, both current and anticipated that reflect a marked departure from some earlier file system assumptions. This has led to reexamine traditional choices and explore radically different design points. The file system has successfully met different storage needs. It is widely deployed within Google as the storage platform for the generation and processing of data used by our service as well as research and development efforts that require large data sets.

The weighted voting system [7] in p2p explains a decentralized mechanism for providing sequentially-consistent access to data in a partially connected computing environment. This work has been developed in the context of pervasive computing environments. In pervasive computing scenarios it is common for devices to continually arrive and depart and disconnections can be the rule rather than the exception.

In conservative approach, maintaining strong consistency guarantees and trying to improve availability and performance within those bounds. To achieve these goals and minimize management overhead, a new concept has been developed, a

decentralized weighting-voting algorithm, which guarantees sequential consistency. The algorithm distributes versioned metadata along with the data and allows online reconfiguration by using the same quorums to manage both the data and the metadata.

Tapestry is an overlay location [1] and routing infrastructure that provides location-independent routing of messages directly to the closest copy of an object or service using only point-to-point links and without centralized resources. The routing and directory information within this infrastructure is purely soft state and easily repaired. Tapestry is self-administering, fault tolerant, and resilient under load.

III. EXISTING SYSTEM

Now days a single server has the capability to handle the multiple requests from the user. But the server has to process the all the requests from the user parallel, so it will lead to a hike in the processing time of the servers. This may leads to loss of data and packets may be delayed and corrupted. On doing this the server cannot process the query from the user in a proper manner. So the processing time gets increased. It may leads to traffic and congestion. To overcome these problems we are going for the concept called “cloud computing”. In this cloud computing we are going to implement the chunk server to avoid these problems.

Google Inc. has a proprietary cloud computing platform which was first developed for the most important application of Google search service and now has extended to other applications. Google cloud computing infrastructure has four systems which are independent of and closely linked to each other.

They are Google File System for distributed file storage, Map Reduce program model for parallel Google applications, Chubby for distributed lock mechanism and Big Table for Google large-scale distributed database.

A GFS cluster consists of a single master and multiple chunk servers and is accessed by multiple clients. Chunk servers store chunks on local disks as Linux files and read or write chunk data specified by a chunk handle and byte range. The master maintains all file system metadata. This includes the namespace, access control information, the mapping from files to chunks, and the current locations of chunks.

When a client wants to visit some data on a chunk server, it will first send a request to the Master, and the master then replies with the corresponding chunk handle and locations of the replicas. The client then sends a request to one of the replicas and fetches the data wanted.

Google File System (GFS) to meet the rapidly growing demands of Google's data processing needs. GFS shares many of the same goals as previous distributed file systems such as performance, scalability, reliability, and availability. However, its design has been driven by key observations of our application workloads and technological environment, both current and anticipated, which reflect a marked departure from some earlier file system design assumptions. GFS client code linked into each application implements the file system API and communicates with the master and chunk servers to read or write data on behalf of the application. Clients interact with the master for metadata operations, but all data-bearing communication goes directly to the chunk servers.

IV. PEER TO PEER STORAGE

It is networking is a method of delivering computer network services in which the participants share a portion of their own resources, such as processing power, disk storage, network bandwidth, printing facilities. Such resources are provided directly to other participants without intermediary network hosts or servers. Peer-to-peer network participants are providers and consumers of network services simultaneously, which contrasts with other service models, such as traditional client-server computing where the clients only consume the server's resources. P2P is a great fit for Cloud storage systems offering the much needed reliability. It provides improved reliability than the client-server cloud.

The storage policies of P2P say that:

- The cost to the P2P system will be lower if one allocates large files to unreliable peers, and assigns smaller files to reliable peers.
- Unreliable peers should be allowed to distribute less, and reliable peers should be allowed to distribute more.
- Smaller files should be assigned a higher distribution cost, and larger files should be assigned with a lower distribution cost.

V. PROPOSED SYSTEM

New cloud storage architecture based on P2P and designs a prototype system. The system based on the new architecture has better scalability and fault tolerance. A cluster consists of a single database and multiple chunk servers and is accessed by multiple clients. Chunk servers store chunks on local disks and read or write chunk data specified by a chunk handle and byte range. The database maintains all file system metadata. This includes the namespace, access control information, the mapping from files to chunks, and the current locations of chunks. When a client wants to visit some data on a chunk server, it will first send a request, and the database then directs with the corresponding chunk handle and locations of the replicas. Hence the processing loads on servers are balanced.

The architecture consists of mainly 3 modules:

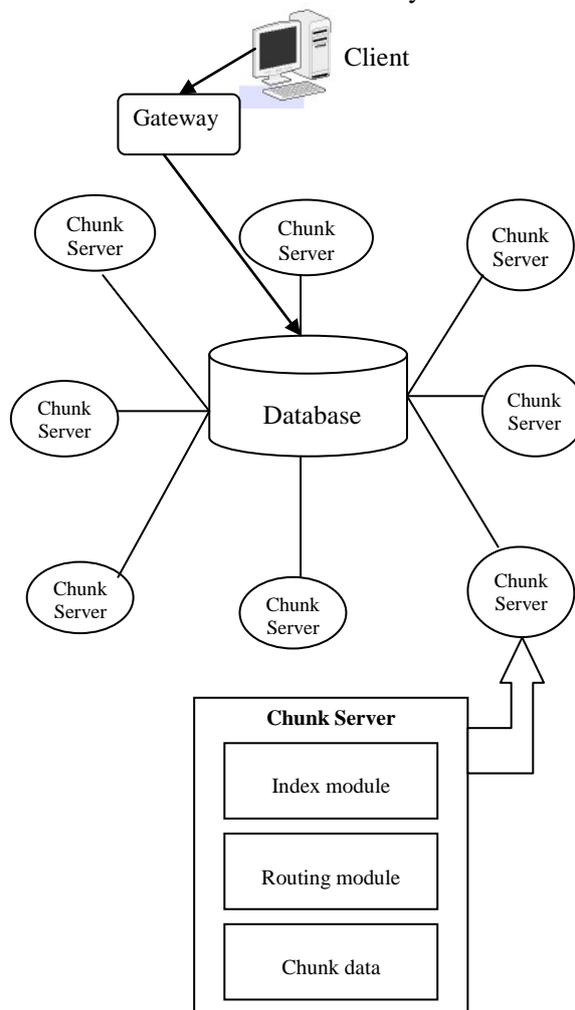


Figure 1: System architecture

Client

The client application is designed to get the data from the platform. Here the client sends user name and password for getting authentication. The authentication for client access is given, if and only if both the user name and password matches to one of the details in database. Else access is denied. After authentication the client send request to gateway. The client gets response after processing the request from gateway. The client can send any request to the server through the gateway. Only registered client can get the service from the server via the gateway. Thus security is implemented in client module also. Client App sends a request for a data block with logic identifier to Gateway.

Gateway

This entity can transfer the request or response between the Client App with the network and can lead the request to the nearest node in the network. This is the important module which acts as an intermediary between the client and the Chunk server. It receives the client's request and forward the request to the nearest chunk server and then it receives the response messages from the chunk server and forward that message to corresponding client/requester. Gateway constructs a P2P search request data package including the logic ID, and sends the request to the chunk server P2P network. Gateway constructs a P2P search request data package including the logic ID, and sends the request to the chunk server P2P network.

Chunk server

This entity is served as the data resource node and P2P node. Different with the function of pure data storage in GFS, the chunk server here has three function modules with separated interfaces. As shown in the figure above: Index Module, take charge of part of the global resource index which is assigned by DHT arithmetic such as Chord, Pastry and so on. Route Module; pass a lookup request by a next hop routing table which is also assigned by DHT. Data Module, provide the data resource stored in the local machine before a Client App can do its work, data blocks and the corresponding replica should be uploaded to the Chunk Servers. How to select the chunk servers for storage is the same with the traditional cloud computing platform.

VII. CONCLUSION AND FUTURE WORK

We propose a cloud computing architecture based on P2P which provide a pure distributed data storage environment without any central entity for controlling the whole processing. The advantage of this is architecture is that it prevents the bottleneck problem that arises in most of the client server communications. The proposed system does its operation based on the performance of the system. It does the monitoring operation to find out the best chunk servers within the P2P network. It does this operation in order to perform efficient resource utilization and load balancing of the servers.

The future work of this proposed system could to modify the system performance by reducing the number servers present in the network. It a tough job to manage a lot number of servers. The enhancement says that if the operation is performed, for example, with the help of 100 servers, then reduce the number of servers to 50 servers by increasing the capacity of each server. Then pipelining concept could also be introduced within this P2P network in order to provide faster access. By enabling all these concepts the architecture provides better scalability, manageability, fault tolerance, better performance.

VIII. REFERENCES

- [1]Ben Y. Zhao, John Kubiatowicz, and Anthony Joseph, "Tapestry: An Infrastructure for Fault-tolerant Wide-area Location and Routing", Technical Report No. UCB/CSD -01-1141, University of California Berkeley.
- [2]Francesco Maria Aymerich, Gianni Fenu, Simone Surcis. An Approach to a Cloud Computing Network. 978-424426249/08/\$25.00 ©2008 IEEE conference.
- [3]Boss G, Malladi P, Quan D, Legregni L, Hall H. Cloud computing. IBM White Paper, 2007. http://download.boulder.ibm.com/ibmdl/pub/software/dw/wes/hipods/Cloud_computing_w_p_final_8Oct.pdf.
- [4]Ghemawat S, Gobioff H, Leung ST. The Google file system. In: Proc. of the 19th ACM

Symp. On Operating Systems Principles. New York: ACM Press, 2003. 29_43.

[5] Ke Xu, Meina Song, Xiaoqi Zhang, Junde Song, "A Cloud Computing Platform Based on P2P".

[6] Burrows M. The chubby lock service for loosely-coupled distributed systems. In: Proc. of the 7th USENIX Symp. on Operating Systems Design and Implementation. Berkeley: USENIX Association, 2006. 335_350.

[7] Oasis: M. Rodrig, and A. Lamarca, "Decentralized Weighted Voting for P2P Data Management," in Proc. of the 3rd ACM International Workshop on Data Engineering for Wireless and Mobile Access, pp.85-92, 2003.

[8] Antony Rowstron and Peter Druschel, "Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems", IFIP/ACM International Conference on Distributed Systems Platforms.

[9] 14] Sigma: S. Lin, Q. Lian, M. Chen, and Z. Zhang, "A practical distributed mutual exclusion protocol in dynamic peer-to-peer systems," in Proc. of 3rd International Workshop on Peer-to-Peer Systems (IPTPS'04), 2004.

Assessment of environmental impact of Rain Catching and Controllable Irrigation regime in paddy field for sustainable agriculture in Nanjing, China.

Guo Xiangping, Bougourou Soulemane, Zhang Zhanyu

Abstract— In a healthy farm system, agriculture works in harmony with the natural environment. This begins with healthy soil that stores water, nutrients and provides a stable base to support plant roots. In a sustainable agriculture system, soil is kept in balance. To assess environmental impacts of Rain Catching and Controllable Irrigation on a paddy field, a study on the experimental field was performed during the growing season over a 2-year period. The plots were separated according to controllable irrigation schedule T2 (High dry low flooding) and T3 (High dry High flooding) rather than conventional irrigation regime T1 (Shallow and frequent irrigation). The mechanism of RCCI model showed its ability to reduce the water supply irrigation by 36% for T3 and 21% for T2, while T1 treatment provided to the rice plants 100% of its water requirements. The maximum use of rainfall by reducing surface drainage and percolation on the plots was the issue of the RCCI model. The results showed that T3 treatment got roots highest activities (285 $\mu\text{g/g.h}$), T2 treatment take the medium level of roots activities (247.26 $\mu\text{g/g.h}$), whereas T1 was the last one (226.66 $\mu\text{g/g.h}$) during the same rice growth period. The T3 treatment had present the half of nitrogen lost (9.17kg/ha) of the T1 treatment (20.28kg/ha). The RCCI model also reduces at least half phosphorus losses by reducing the volume of drainage water from 150.25mm (T1) to 84.14mm (T3). T3 treatment had a higher actual rice grain yield (7.56 T/Ha), and was a beneficial treatment with less environmental pollution. Pests in the paddy field were more important in 2011 than 2010. The weeds *Echinochloa pyramidalis* increased from 4plants/m² to 9.5 plants/ m². The rice yellow stem borer, *Scirpophaga incertulas* attacks were earlier and sterner in 2011 and caused a huge economic loss P= 10.39%.

Keywords — controllable irrigation, root activities, nitrogen and phosphorus loss, weeds *Echinochloa pyramidalis*, yellow stem borer *Scirpophaga incertulas*, rice yield loss.

◆

1 Introduction

United Nation's predictions of global population increase for the year 2025 require an expansion of food production of about 40-45%. Irrigation agriculture will be an essential component of any strategy to increase the global food supply[1]. Without irrigation, increases in agricultural yields and outputs that have fed the world's growing population would not have been possible[2]. The number of irrigated land is constantly increasing with the development of various technologies through the world. Asia holds the bulk of the irrigated land with 37 percent of the land under cultivation in the region[3]. This is the highest level compared to the other major regions of the world. Democratic People's Republic of Korea has the highest level, with 73 percent of cultivated land under irrigation, followed by Japan with 65 percent and China with 55 percent (545,960 square kilometers). Most of China is unproductive agriculturally. Arable land is concentrated in a band of river valleys and along the southern and eastern coasts [4]. These rivers and valleys have been polluted by pesticide residues and fertilizers led by drainage water. Agriculture is the most source of pollution [5]; by Nutrients (phosphorus or nitrogen)[6] and pesticides [7],[8]. Reducing this pollution becomes a national concern.

Continuously rapid growth of domestic and industrial water uses, growing recognition of environmental demands for water, and the high cost of developing new water resources

threaten the availability of irrigation water to meet growing food demands [2]. Because of that, each country has improved its irrigation techniques to reduce water supply to the crops without affecting their performance. So in China since the 1980s, the efficient irrigation regimes for rice have been researched, and many of those have been adopted in different rice growing regions, aiming to increase the water and land productivity [9], [10]. Also expansion of the irrigated area should be limited if water resources are to be conserved and the natural environment protected. In order to ensure the sustainable development of agriculture, especially food security, the future water-resources strategy must focus on changes in the agricultural water-saving technology to increase the effective use of precipitation and irrigation water[11]. Environmental considerations suggest that irrigation water supply for Chinese agriculture should be maintained at around 320–340 billion m³ a year. However, the state has suggested that the country must produce another 50 million tons of grain per annum by the year 2020. This suggests an increasing water requirement for agriculture [12], whereas the climatic change disrupts the water regimes. The threat has led researchers to develop a large number of irrigation practices using less water such as Rain Catching and Controllable Irrigation, without considering the environmental consequences. Therefore, experimental evidence is still not reported in the

international literature on the environmental impact of that irrigation regime adopted since five years.

The objective of the current study is to focus on the assessment of the environmental influence of Rain Catching and Controllable Irrigation model which has been in use (since 2006) in paddy fields in Nanjing, China. Its aim is to review different aspects of the environmental impacts of RCCI in areas relating to soil conditions improvement proven by root activities, such as total nitrogen and phosphorus loss in drainage water, and also on rice pests.

2 Materials and Methods

Experiment site and field soil conditions

The experiment was conducted from June to October 2010 and June to October 2011 at the Water Saving Park Agricultural Experimental Farm at Soil and Water Engineering Department at Hohai University in Nanjing, China. The farm is located at 31°95'N, 118°83'E, in a suburb of Nanjing at an area downstream of the Yangtze River drainage basin with an average elevation of 15 m above the sea level[13]. This area is characterized by a humid subtropical climate and is under the influence of the East Asia Monsoon. The mean annual temperature is 15.5 °C, with monthly mean ranging from 2.4 to 27.8 °C; the highest temperature in this area is 43.0 °C while the lowest is -16.9 °C. The average annual rainfall is 1062 mm. The soil at the experimental site is clayey loam (33.81%) clay (65%) silt 0.22%, and 0.97% (sand) with a pH (H₂O) of 8.06 and field capacity of 29.3; Table 2.1 shows the physical and chemical properties of the field soil.

Table 2.1: Soil physical and chemical properties

Soil texture	Clay
pH	8.06
Organic matter (mg/kg)	12.26
Soil depth (Cm)	0-20
Total phosphorus (mg/kg)	330.9
Available P (mg/kg)	10.13
N Total (%)	0.1
Available nitrogen (mg/kg)	65

During the growth season rainfall in 2010 from June to September was 580mm and the number of rainfall days for growth season of 2010 was 43days. During the growth season rainfall 2011 from June to September was 601mm and

the number of rainfall days for growth season of 2011 was 54 days. The average temperature and humidity during the growing season in this area are 30°C and 79.75% respectively [14]

At this experiment, 1-month-old rice seedlings (*Oryza sativa* cv. Nanjing 44) were transplanted in June and harvested in September during 2010–2011. A fertilizer rate of 55, 45, 40, 10, 3 kg/ ha of N, P, K, S, Zn in the form of triple super phosphate, muriate of potash, gypsum and zinc sulphate, respectively were applied as basal dose at final land preparation following the local farming practices without spraying Insecticides and Herbicides.

Experimental design and treatments

The experimental design was based on the new concept of "Rain Catching and Controllable Irrigation (RCCI)" of rice[15]. Details of the design are shown in Table 2.2.

Table 2.2: Experimental design of Rain Catching and Controllable Irrigation (unit: mm)

Treatments	RCCI model		TRI model
	T3	T2	T1
Growth period	High dry High flooding	High dry Low flooding	Shallow and frequent irrigation
Seedling	10-30-70	10-30-70	10-30-70
Tillering	Early - Tillering	80%-100%-80	0-30-70
	Middle	70%-100%-120	0-30-90
Stem elongation	Late	70%-100%-100	0-30-120
		80%-100%-200	0-30-120
Heading	80%-100%-200	80%-100%-150	0-30-100
Milky stage	80%-100%-80	80%-100%-80	0-30-60
Ripening period	70%-80%	70%-80%	70%-80%

Note: (1) The three data, for example 10-30-70 respectively means the lower limit of irrigation, upper limit of irrigation and the maximum water-catching depth after rain in Table above; (2) Percentage means the percentage of the average moisture content of field capacity water content in the upper 30cm of soil and other units are mm. RCCI model: Rainfall Catching and Controllable Irrigation model; TRI model: Traditional Irrigation model or Conventional Irrigation schedule.

The experiment was conducted on the natural vegetation, which consisted of: *Poaceae* (*Echinochloa pyramidalis*, *Bromus sp.*, *Dactylis glomerata*, *Digitaria ischaemum*...); *Typhaceae* (*Typha australis*, *T. latifolia*, *T. albida*, *T. alekseevii*, *T. angustifolia*); *Fabaceae* comprises three subfamilies (with distribution and some representative species): *Mimosoideae*, *Caesalpinioideae*, *Faboideae*; *convulvaceae* [16], [17].

Each plot measured 8 m long and 2 m wide, and was repeated four times in a completely random block design, the plants were transplanted on a scale of 0.2 m x 0.2 m giving a density of 250,000 plants per hectare with two

plants per hill. In each plot, there were inserted pots; these pots had 80 cm of diameter and 60 cm of height with a content of 4 reference plants. The treatments applied on the plots were also applied to the pots. For this experiment, 1-month-old rice seedlings (*Oryza sativa* cv. Nanjing 44) were transplanted in May and harvested in September during 2010–2011 summer. Cultivation regimes were consistent with optimum rice production in the region.

Water depth and soil moisture measurement

Soil moisture was measured at 0- 30 cm of soil profiles for unsaturated by Time Domain Reflectometer (TDR) and water depth in the field was measured by a ruler every 2-4 days interval.

Counting of weeds

Spatial distribution of weeds is characterized by weed density data collected at locations in a field (Density: Is a measure of abundance per unit area). The total number of weeds for all species was counted by adding emerging species every 5 days. Emerging species are a newly established weed species whose extent, distribution and abundance is expanding (trend is increasing), and whose impacts are likely to be significant.

Absolute estimates of insects

Absolute estimates of the actual insect density are counted directly on the plant in time and space. An absolute estimate can be defined as a count of insect numbers with reference to a predefined unit of measurement. The count per unit measurement provides an estimate of insect density and can be recorded in terms of an unit area, plant or plant part, e.g. numbers of eggs per leaf, the number of larvae per plant, and the number of pupae per square meter [18].

Laboratory procedures of Phosphorus and Nitrogen determination

In the key Laboratory of Efficient Irrigation-Drainage and Agricultural Soil-Water Environment in Southern China, Hohai University, drainage water samples were analyzed by the standard methods of APHA (1995) [19] for total nitrogen (TN) and total phosphorus (TP) concentrations.

Estimated Mass loss = drainage volume × concentration (2.1)

Determination of rice roots activities

White roots were cut and 0.5g of the material (roots) was dried with a blotting paper. It was then put in a test tube and 5 mL of 0.4% Triphenyltetrazolium Chloride TTC ($C_{19}H_{15}N_4Cl$) + 5 mL of (Buffer solution NaH_2PO_4 / Na_2HPO_4) phosphate buffer at pH 7.0 added; the mixture was incubated at 37°C in water bath for 1 hour. The reaction was stopped by adding 2 mL of 1 mol/L H_2SO_4 . The roots after drying it with the blotting paper, was grinded in the mortar with 3 to 4 mL of Ethyl acetate ($C_4H_8O_2$) and adding Ethyl

acetate ($C_4H_8O_2$) to get 10 mL. The samples were then centrifuged at 4000 revolutions per minute (r.p.m) into the sterilization machine (Anke TDL80-2B) for 4 min. Use 10 mL of Ethyl acetate ($C_4H_8O_2$) in one test tube for the control (2 ck) and the absorbance of the supernatant measured at A_{485} nm in the spectrophotometer. The Formula below was used for the calculations: Triphenyl Tetrazolium Formazane (TTF).

$$TTF (\mu g) = 789.45 A_{485} + 7.3712.$$

Roots Activities ($\mu g/g.h$) = TTF/ sample fresh weight.

Water Use Efficiency (WUE)

According to Barrett Purcell & Associates (1999) [20] instead of Water Use efficiency it is in fact better to use the term Water Use Indices.

Evapotranspiration and Crop Water Use Indices (WUI)

ET (Evapotranspiration) was measured by the plastic pots buried in each plot. The size of the pots was 80 cm in diameters and 60 cm in height. The plant density kept same as that in field. The moisture was kept the same as that of the field.

Total water consumption (W_c) of rice was measured by difference of water depth or water moisture (if there was no surface water in the field or in the pots of the top soil 30 cm). Percolation can be calculated as follows:

$$P_1 = W_c - P - I + R + D \quad (2.2)$$

Where P is precipitation (mm), I is irrigation water amount (mm), R is runoff/ run-on (mm), Runoff was considered zero because the experimental plots were surrounded with dikes and D is surface drainage (mm) amount from paddy field. P_1 (mm/day) is the percolation amount of water from the root zone.

Irrigation water use indices (WUI)

$$WUI \text{ (Kg/ML)} = \frac{\text{yield}}{\text{Irrigation water applied}} \quad (2.3)$$

$$\text{Crop WUI (Kg/mm)} = \frac{\text{yield}}{ET} \quad (2.4)$$

Where: ML: Mega Liter = 1 000 m³, ET=Evapotranspiration

Calculation of crop yield loss

Yield loss modeling is based on a set of concepts that were developed within the last two decades by FAO (2005) [21] on production ecology and plant protection. The main principles of this method are:

$$C = \text{nuisance factor: } C = (a - b) * 100/a \quad (2.5)$$

Error! Reference source not found. , and

$P = \frac{C \cdot I_p}{100} \cdot a$ (2.6)
where a = average yield per non infested plant, b = average yield per infested plant and I_p = percentage of infest plants.

Statistical Analysis

Treatment effects in the experiment were analyzed through using analysis of variance (ANOVA) procedure of SPSS software Version 14.0. Average treatment were separated by Least Significant difference (LSD) test at $p \leq 0.05$ unless specified. Also, Excel and Origin8 were used for data analysis as well as plotting graphs and figures.

3 Results and discussions

Mechanism of Rainfall Catching and Controllable Irrigation in paddy field Rice Evapotranspiration and Percolation quota reduction of RCCI

The Table 3.1 below shows Evapotranspiration and Percolation rate under RCCI model. The RCCI model has reduced the evapotranspiration from 37.73% by T3 and T1 treatment to 25% by T2 and T1 treatment. These results confirm those of Mao and Cui (2001) which show a reduction of Evapotranspiration from 5 to 30% of controllable irrigation. Under RCCI model, most of the time the average soil moisture content in the rice root zone (0-30cm) was in 70%-80% of field capacity. This leads to the average soil moisture content in 0-5cm of the surface layer to the level of below 50% of field capacity. Under this condition, the rice growth is not affected but the evaporation from the soil in paddy field can be reduced by about 10-20%[22].

Table3.1: Evapotranspiration and Percolation quota in the paddy fields

Treatments	ET(mm)	I (mm)	D (mm)	P1 (mm)	Effective rainfall(mm)
T1	672.21a	603.6a	150.25a	676.2	450.75
T2	504.21b	480.5b	102.17b	673.05	498.83
T3	418.58c	386.4c	84.14c	646.8	516.86

In the column, averages followed by the common letter(s) are not significantly different at level of $P \leq 5\%$.

The reduction of percolation of RCCI model results from two ways: (1) The duration of no water depth and unsaturated condition in the paddy field is longer under RCCI than that under conventional irrigation schedule (T1); (2) The depth of water depth is shallower under RCCI than that under Traditional Irrigation. Under RCCI, the percolation was

reduced due to the above-mentioned two conditions; Table 3.1 confirms that with T3 treatment which had the lowest (646.8mm) percolation rate P1.

Increasing rainfall utilization of RCCI

The intent of Rainfall Catching and Controllable Irrigation (RCCI) concept is to reduce as much as possible water supply for irrigation while using the maximum rainfall (Table 3.1). It means that irrigation will come in complement to avoid the water stress to the plants. Though the lower limits of RCCI are similar to those of T1, the rain-catching depth of RCCI was much higher than those of conventional irrigation. The capacity of paddy fields to store rainfall is increased greatly, and precipitation is fully utilized without hindering rice growth under RCCI.

Table3.2: Irrigation quota in the paddy fields unit: mm

Treatments	T1	T2	T3
Seedling	80.32a	67.5b	60.06c
Tillering	100.4a	90.5b	71.28c
Elongation	120.72a	92.6b	77.28c
Heading	120.72a	98.4b	77.28c
Milk stage	181.44a	131.5b	100.5c
Amount Irrigation	603.6a	480.5b	386.4c
Irrigation Schedule (Times)	14a	10b	10b

In the row, averages followed by the common letter(s) are not significantly different at level of $P \leq 5\%$.

According to the treatments applied, the least irrigation water delivery in the field that data revealed in T3 which (386.4mm) has high control over rainfall water storage; followed by T2 (480.5mm) and T1 got the highest irrigation water delivery in the field (603.6mm) according to the controllable irrigation regime schedule as shown in Table 3.2.

Water Use Efficiency by RCCI model

The Crop Water Use Indices (Crop WUI) for the T3 treatment (18.06Kg/mm) was the highest whereas T1 (control) produce only 9.97 Kg/mm. T3 treatment used 1Mega Liter of irrigation water to produce rice grain yields of 1.95 tons, whereas the control T1 used the same amount to produce 1.11 tons. The RCCI model used efficiently water input (rain and irrigation) for grain production, which is significantly different ($P \leq 5\%$) from the conventional irrigation. Under T3 treatment crop productivity was the highest and most efficient water use.

Table3.3: Crop Water Use Indices

Treatments	T1	T2	T3
------------	----	----	----

Yield (Kg/Ha)	6700a	6630a	7560b
Irrigation (ML)	6.036a	4.803b	3.864c
ET (mm)	672.21a	504.21b	418.58c
Irr. WUI (Kg/ML)	1110.01a	1380.39b	1956.52c
Crop WUI (Kg/mm)	9.97a	13.15b	18.06c

In the row, averages followed by the common letter(s) are not significantly different at level of P≤5%.

The causes of the reduction of irrigation water requirements by using RCCI are that the percolation and evapotranspiration decreased remarkably while the utilization of rainfall increased as shown in Table 3.1. T1 (conventional irrigation) treatment provided 100% of rice plant water requirements while T2 and T3 (RCCI) treatments provided only 79% and 64%, respectively, without reducing the grain yield. RCCI had saved 215.07 mm of water for the T3 treatment and 120.97 mm for T2 treatment.

Treatment T3 used 554L of input water to produce 1kg of rice grain while T2 treatment used 760L. According to Zwart and Bastiaansen (2004), an average water of 2,500 liters needs to be supplied (by rainfall and/or irrigation) to a rice field to produce 1 kg of rough rice. These 2,500 liters account for all the outflows of evapotranspiration, seepage, and percolation. The average number is derived from a large number of experimental data at the individual field level across Asia [23]. So, RCCI model saved 1946L and 1740L for treatment T3 and treatment T2, respectively.

Rice production must be viewed in the light of the emerging water crisis, as climate-change-induced shifts in rainfall patterns combined with the diversion of irrigation water for urban and industrial uses.

Environment protection of RCCI technic

RCCI Improves soil conditions in paddy field

Root activities during the main steps of rice cycle are used to assess soil conditions under the RCCI model and conventional irrigation. Results showed that roots activities increased with the development of rice plants (Table 3.4) under Controllable Irrigation conditions and declined with ageing plants. Rice roots activities were high in T3 (High dry High flooding) in every rice growth stage, from 225.86 μg/g.h at tillering stage, 285 μg/g.h at elongation stage to 54.05 μg/g.h at rice milk stage. Under treatment T1 roots activities are lowest in almost all growth stages of rice from (184.22 μg/g.h) tillering stage to 39.38 μg/g.h at milk stage. Treatment T2 got the medium level of roots activities during the same period. Treatment T3 had the average longest roots 29.50cm and treatment T1 had the average shortest roots 21.25cm. In waterlogged soil (T1), diffusion of gases through soil pores was so strongly inhibited by their water content that it fails to match the needs of roots growing. A slowing of oxygen influx is the principal cause of an

injury to roots, and the shoots they support [24]. The maximum amount of oxygen dissolved in the floodwater in equilibrium with the air is a little over 3 % of that in a similar volume of air itself. This small amount of oxygen is quickly consumed during the early stages of flooding by aerobic micro-organisms and roots. In addition to imposing oxygen shortage, flooding also impedes the diffusive escape and/or oxidative breakdown of gases such as ethylene [25] or carbon dioxide that is produced by roots and soil micro-organisms. This leads to accumulations that can influence root growth and function. Traditionally, a proper rate of deep percolation is maintained to leach the poisonous matters within rice root zone, resulting from an anaerobic condition and bring oxygen into rice root zone. The longer the duration of the soil submerged by deep water, the lower is the content of dissolved oxygen in soil water. These hydromorphic conditions reduce significantly (P≤5%) the rice roots activities as shown in Table 3.4

Table 3.4: Rice roots activities under Controllable Irrigation conditions

Growth stages	Treatments	Root length (cm)	Roots activities (μg/g.h)
Tillering stage (35DAT)	T1	20	184.22a
	T2	21.25	209.88b
	T3	22	225.86c
Elongation (50DAT)	T1	21.25	226.66a
	T2	21.75	247.26b
	T3	23.25	285c
Heading (65DAT)	T1	21.70a	88.23a
	T2	21.81a	133.83b
	T3	25.25b	152.51c
Milk stage (80DAT)	T1	21.25a	39.38a
	T2	21.75a	39.77a
	T3	29.50b	54.05b

In the column, averages followed by the common letter(s) are not significantly different at level of P≤5%.

The action of micro-organisms can be promoted and the accumulation of poisonous substances in the soil can be avoided by the favorable soil aeration. The microorganisms in the soil under RCCI model are more abundant than those under conventional irrigation. The soil fertility can be increased through the transformation of organic matter by the abundance of important microorganisms. The rice roots grow well under oxidized paddy fields even under moderate water stress, 29.5cm for T3 treatment was a proof.

Drainage and minerals (Nitrogen and Phosphorus) lost under RCCI model

The results in Table 3.5 showed the amount of drainage water during the whole growing period of rice. The smallest amount was observed in T3 plot 84.14 mm whereas T1 plot showed the largest amount 150.25 mm. The drainage water during the seedling period was less than the other phases of rice development for all the treatments. Tillering period for all the treatments had the biggest drainage water amount.

Table3.5 : Drainage in the paddy fields unit: mm

Growing period	T1	T2	T3	Total
Tillering	50a	34b	42.07c	126.07
Stem elongation	50.16	34.1	-	84.26
Heading	-	-	42.07	42.07
Milk stage	50.08a	34.05b	-	84.13
Amount of Surface drainage	150.25a	102.17b	84.14c	336.56
Frequency of drainage (Times)	3	3	2	-

In the row, averages followed by the common letter(s) are not significantly different at level of P≤5%.

Most of nitrogen loss is associated with the combination of excessively wet soil, the results in Table 3.6 show the estimated mass of total nitrogen loss in drainage water volume. Total nitrogen is composed of three forms of nitrogen, which are mainly found in soil drainage solution, namely, NH₄-N, NO₂-N and NO₃-N; that estimated mass of total nitrogen loss was increased gradually and doubled with the drainage volume, from T3 (9.17 kg/ha) to T1 (20.28 kg/ha). The RCCI model by reducing the volume of drainage water reduces also total nitrogen loss.

Table3.6 : Total Nitrogen loss in paddy fields

Treatments	Average concentration (mg/L)	Total drainage Volume (m3/ha)	Estimated mass of T N loss (kg/ha)
T3	10.9	841.4a	9.17a
T2	13.7	1021.7b	13.99b
T1	13.5	1502.5c	20.28c

In the column, averages followed by the common letter(s) are not significantly different at level of P≤5%.

The results in Table 3.7 below show the estimated mass of phosphorus loss under these three treatments. The control T1 had lost four times (0.648kg/ha) phosphorus more than T3 (0.149 kg/ha). While the phosphorus loss under the T2 treatment (0.32 kg/ha) was half of that cause by T1 treatment. The RCCI model reduces at least half phosphorus losses by reducing both volume of drainage water and

concentration of TP.

Table 3. 7 : Phosphorus loss in paddy fields

Treatments	Average concentration (mg/L)	Total drainage Volume (m3/ha)	Estimated mass of P loss (g/ha)
T3	0.18	841.4a	151.45a
T2	0.32	1021.7b	326.94 b
T1	0.43	1502.5c	646.08c

In the column, averages followed by the common letter(s) are not significantly different at level of P≤5%.

According to Weining, (1993), alternate flooding and drying can reduce 20% to 65% of the percolation and seepage water from rice fields [26] and Fertilizer loss is brought about by this way of infiltration and drainage. This confirms the results of Table: 3.6 and Table: 3.7 that show the reduction of total nitrogen and phosphorus loss in the drainage water by the RCCI treatment regarding to the conventional irrigation. The higher rain-catching depth under RCCI weakened kinetic energy of raindrops thus decreased turbulence of surface water. That prevented the amount of topsoil particles rich in particulate N and P as well as soluble N and P, to enter into surface water. By this way RCCI reduced soil erosion. Additionally, the increasing of rain-catching depth also prolonged residence time of rain water in paddy field, and thus promoted deposition of soil particles, absorption of plants and soil, as well as nitrous nitrification and denitrification, which could reduce concentrations of TN and TP in surface water.

Pests in paddy field under Controllable irrigation regime

Weeds in the paddy fields

Weeds are common in transplanted wetland rice and they are highly competitive to the crop [27]. The occurrence of weeds has become a serious problem and they limit the yield and quality of crops. It is often stated that some weeds cause total crop failure and that weeding practices are absolutely essential [28], [29]. Optimum yields can be obtained only when the crop is free from weeds. Consequently, weed control has always been a major input in rice production. Under alternation of drying and flooding conditions, some species of weeds have been emergent, the hydromorphic conditions with more or less water layer reveal another kind of weeds.

Several genuses and species of weeds have been identified in plots; Table 3.8 and Table 3.9 give the detail. However, in 2011, we focused our attention on *Echinochloa pyramidalis* species.

Table3.8: Major rice weeds in the paddy fields (2010)

Weed species	Density in different plots (plant/m2)		
	T1	T2	T3

<i>Echinochloa pyramidalis</i>	4.0a	1c	1
<i>Cyperus spp</i>	2.0b	0.4d	0.7
<i>Commelina diffusa</i>	1c	0.2d	0.4
<i>Marsilea quadrifolia</i>	1c	0.2d	0.5

In the column, averages followed by the common letter(s) are not significantly different at level of P≤5%.

Table 3.9 : Major rice weeds in the paddy fields (2011)

Weed species	Density in different plots (plant/m ²)		
	T1	T2	T3
<i>Digitaria ischaemum</i>	1a	1a	4a
<i>Cyperus difformis</i>	3b	1a	0.7b
<i>Commelina diffusa</i>	1a	1a	1b
<i>Marsilea quadrifolia</i>	1a	0.2b	0.5b
<i>Dactylis glomerata</i>	2ab	2c	1b
<i>Polygonum lapathifolium</i>	1a	1a	1b
<i>Alamo rental</i>	1a	1a	1b

In the column, averages followed by the common letter(s) are not significantly different at level of P≤5%.

One reason for flooding rice is to manage a broad spectrum of terrestrial weed species that are sensitive to flooding. Flooding effectively controls many problematic weed species [30]. Total weed density and number of weed species were higher in 2011 (Table 3.10) than in 2010 (Table 3.9). It appears that sufficient accumulation of surface water in paddy fields can prevent germination and growth of many weeds under the treatments T1 and T2. The dominant weed species in the field can be regrouped in two types as shown in the Table 3.9, the Poaceae and the Cyperaceae. The Poaceae (*Digitaria ischaemum* 4 plants/m²) dominate the drier plots T3 treatment, the Cyperaceae (*Cyperus difformis*), with 3 plants/m² are abundant on the wetter plots T1 treatment. Also it is notable that the species found in both species and the two conditions (*Commelina diffusa*, *Polygonum lapathifolium* and *Alisma orientale*) of all treatments.

Insect in the paddy field

The figure 3.1 below shows the number of rice plants infested by yellow stem borer *Scirpophaga incertulas* in 2010. The attacks of (*Scirpophaga incertulas*) were observed

from the 60 DAT and reach the peak in the 100 DAT with a maximum of 45 plants counted in the treatments T3.

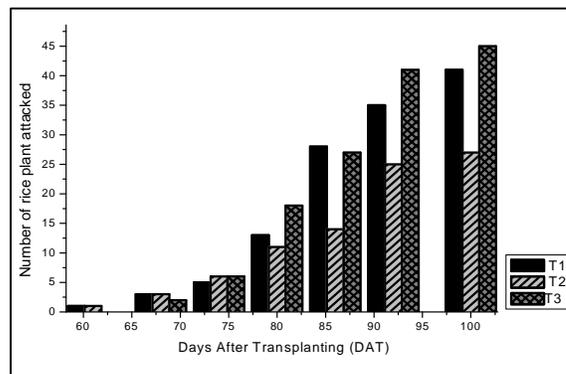


Figure3. 1: Number of rice plant attacked (2010)

The Figure 3.2 by tale makes a link between the density of the weed *Echinochloa pyramidalis* and the infestations of the rice plants by the yellow stem borer (*Scirpophaga incertulas*) in 2011.

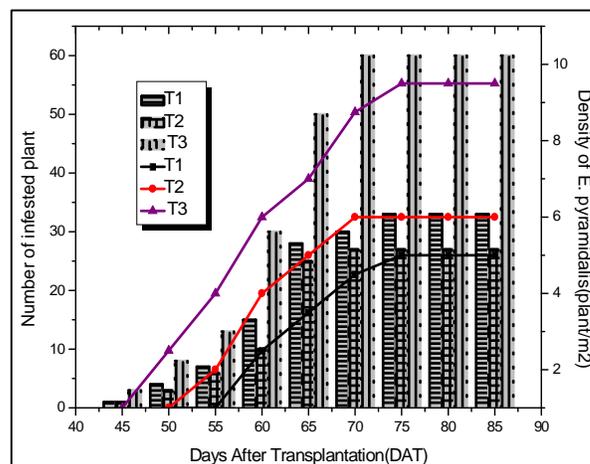


Figure3. 1: Number of rice plant attacked 2011and *E. pyramidalis* density

Figure 3.2 showed first that all treatments have been infested by the weed *E. pyramidalis*, but in different densities (P ≤5%) during the development cycle of the rice. The weed infestation began early from 45 DAT and the peak was

observed at 75 DAT with an average of 9.5 plants per meter square for T3 contrary to 5 and 5.25 plants per square meter respectively for T1 and T2 (see the curves in the Figure 3.2).

The extent of the weeds in 2011 was due to the fact that no weed management was practiced on the plots during the experiment. Weather conditions (rainfall, humidity and temperature were higher in 2011) also contributed to this emergence of those weeds. Weeds especially the graminoid compete more with cereals because of their similar growth behavior, rooting profile and nutrient requirements. Weeds absorb nutrients growing more efficiently than crop.

The histograms in the Figure 3.2 by tale show the number of rice plants infested by the yellow stem borer (*Scirpophaga incertulas*) in 2011. The higher the density of *E. pyramidalis* is raised on the plots, the more serious the borers attack the rice plants. The maximum density of *E. pyramidalis* on the T1 plot was 5plants/m² and the maximum average number of parasitized rice plants in these parcels is about 27. On the T2 plot there were 5.25plants/m² of *E. pyramidalis* and 33/m² infested rice plants. Additionally, on T3 plot there were 9.5plants/m² of *E. pyramidalis* and 60 plants/m² infested rice plants. The T3 treatments had the highest average weeds density (9.5plants/m²) and the highest average infested rice plants (60).

Insect-plant interaction

Insect-plant interaction refers to the activities of two types of organisms: insects that seek out and utilize plants for food, shelter, and/or egg-laying sites, and the plants that provide those resources. The Figure 3.2 shows two kinds of relationships: (1) Link weeds (*E. pyramidalis*) / insects (*S. incertulas*), (2) Link weed density / the number of rice plant attacked.

Weeds can harbor pests and diseases which transfer to the crop [31]. The results showed that *E. pyramidalis* was the medium host plant of rice yellow stem borer *Scirpophaga incertulas*. This results confirms those phenomena [32]. This pyral whose first generation emerged from *Echinochloa pyramidalis* will lay its eggs on rice plants in favor of environmental conditions. The weeding of *E. pyramidalis* is particularly difficult because it germinates with rice plants and emerges in the same hill with the crop. Consequently the insects that *E. pyramidalis* harbors can easily transfer to the rice plants.

Nuisance Factor of Insects (*Scirpophaga incertulas*)

The nuisance of an insect on a plant is the amount of damage that the bug can cause on its host. In farming an insect may be classified as a pest if the damage it causes to a crop is sufficient to reduce the yield and/or quality of the 'harvested product' by an amount that is unacceptable to the farmer. With a rate of 14.81% of infested rice plants (**lp**), the yellow borer *Scirpophaga incertulas* had caused an economic loss of P= 10.39% for the treatment T3. *Scirpophaga incertulas* had induced the highest rate of nuisance (C=0.7) to the treatment T3. On the other hand, the treatment T1 presented the lowest economic loss P=4.33% and the rate of nuisance

(C=0.52) shown in Table 3.10.

Table 3.10 : Economic loss due to yellow bore (*Scirpophaga incertulas*)

Treatments	C	lp (%)	P (%)
T1	0.52	8.25a	4.33a
T2	0.56	7.87a	4.37a
T3	0.70	14.81b	10.39b

In the column, averages followed by the common letter(s) are not significantly different at level of P≤5%.

The damage symptoms of *Scirpophaga incertulas* vary according to the stages of growth of the rice plants. During the very early stages of growth, the larva damaged the growing point in the terminal shoot. This condition is known as 'dead heart'. If the borers attack occurred at the flowering stage, the resulting panicles would become white and empty, known as the 'white head'. The empty paddies do not have any economic value.

Grain yield under controllable irrigation regime

The grain yield is the amount of grain harvested per unit area for a given time (Rabbinge, 1993). In agriculture, the crop yield is a measure of the grains or dry matter quantity in a particular area. It is usually expressed in kilograms per hectare (or metric tons per hectare). The Table 3.11 below points out the results of rice grain yield at 14% of humidity obtained in a field after drying. The lowest grain yield is observed in the control T1 plot (6.7T/ha) whereas the T3 plot showed the highest grain yield (7.56 T/ha). However, the treatments did not affect the rice ear length. The gap between the theoretical yield and the real or actual yield was higher on T1 (3.19 T/Ha) plot than the one under RCCI model treatments.

The resulting yield, obtained in a field injured by one or several pests, is defined as the actual yield (Rabbinge, 1993); it is the yield actually harvested in a farmer's field. Yield loss or damage represents the difference between the attainable and the actual yield, that is, the yield loss caused by pest injuries.

Table 3.11 : Grain yield under controllable irrigation regime

Treatments	Theoretical Yield (T/Ha)	Actual Yield (T/Ha)	Yield gap (T/Ha)	Ear length (cm)
T1	9.89a	6.7a	3.19	17.0
T2	9.25b	6.63a	2.62	17.5
T3	10.35c	7.56b	2.79	17.03

In the column, averages followed by the common letter(s) are not significantly different at level of P≤5%.

The yield gap between the theoretical yield and the real or

actual yield was the result of several combined factors: (1) Yield loss due to weeds that can reach 79% for the density of 269 plants per square meter of weeds; (2) Yield loss due to insect attacks *S. incertulas* that accounts for 14.81% of rice plants causing an economic loss of P= 10.39% can reach 20% according to [34]; (3) Loss of yield due to poor agricultural practices.

4 Conclusion

This study assesses the environmental effect of Rain Catching and Controllable Irrigation on paddy field. The mechanism of RCCI model showed its ability to reduce the irrigation water supply by 36% with maximum use of rainfall, by reducing percolation and evapotranspiration the plots. Also, the RCCI model proved its involvement in environmental protection by improving the soil aeration through the development of rice root activities; the development of microorganisms and by reducing significantly the groundwater pollution and weakening erosion.

Pests in the paddy field were serious in 2011 than 2010. The weeds *Echinochloa pyramidalis* increased from 4 plants/m² to 9.5 plants/m² on RCCI compared with conventional irrigation model. The borer *Scirpophaga incertulas* attacks were earlier and sterner in 2011 and caused a huge economic loss P= 10.39%.

The RCCI treatment showed a performance in water use indices by: lowest irrigation quota (386.4mm), highest crop Water Use efficiency (18.06kg/mm) and lowest drainage amount (84.14mm). The RCCI had present the half of nitrogen lost (9.17kg/ha) of the conventional irrigation treatment (20.28kg/ha). The RCCI model also reduces at least half phosphorus losses by reducing the volume of drainage water from 150.25mm (T1) to 84.14mm (T3). Also, it had a higher actual rice grain yield (7.56 T/Ha) and the lowest yield gap (2.79T/Ha); thus, it was a beneficial treatment with less environmental pollution.

Acknowledgements

This research is supported by the National Science Foundation of China (No. 51079042 and 51179050) and Qing Lan Project. We would like to express our appreciation to Associate Prof Wang W M, Associate Prof Xing W G and the laboratory group for their assistance in the analysis of the samples and data handling.

References

- [1] FAO., 2002. " (Food and Agriculture Organization of the United Nations), "The salt of the earth: Hazardous for food production. World Food Summit Five Years Later. [Outline/Free Full Text]
- [2] M. W. Rosegrant and S. Shetty, 1994. "Production and income benefits from improved irrigation efficiency: What is the potential?," *Irrigation and Drainage Systems*, vol. 8, pp. 251-270
- [3] S. Siebert, P. Döll, J. Hoogeveen, J.-M. Faures, and K. Frenken, 2006. "The Global Map of Irrigation Areas," in *International Workshop on Global Irrigated Area Mapping*, p. 25. [Abstract/Free Full Text]
- [4] L. Changmin and C. Zhikai, *CNBS (China's National Bureau of Statistics) 1999, China Statistical Year Book 1999. People's Republic of China: China Statistics Press*. Beijing, China: CNBS (China's National Bureau of Statistics), 2001, p. 54. [Abstract/Free Full Text]
- [5] USEPA, 1995. "National Water Quality Inventory: 1994 Report to Congress (Appendixes)," Washington, [Abstract/Free Full Text]
- [6] D. L. Osmond, J. W. Gilliam, and R. O. Evans, 2002. "Riparian Buffers and Controlled Drainage to Reduce Agricultural Nonpoint Source Pollution." North Carolina Agricultural Research Service Technical Bulletin 318, North Carolina State University, Raleigh, NC. [Abstract/Free Full Text]
- [7] D. J. Call Call, Brooke L.T., Kent R.J., Brooke L.T., Kent R.J., Poirer, S.H., Knuth M.L., Shobat, P.J. Slick E.J., 1984. "Toxicity, uptake, and elimination of the herbicides alachlor and dinoseb in freshwater fish," *J. Environ. Qual.*, vol. 13, pp. 493-498.
- [8] J. A. Goodrich, B. W. Lykins, R. Jr., and M. Clark, 1991. "Drinking water from agriculturally contaminated ground water," *J. Environ. Qual.*, vol. 20, pp. 707-717,
- [9] Y.H., Li, Dong B., and Yu F, 1999, "Improved irrigation management of paddy fields for sustainable increases in water productivity," in *Emerging technologies for sustainable land use and water management*, p. 36p. [Abstract/Free Full Text]
- [10] T. P. Bouman, B.A.M., Tuong, 2001. "Field water management to save water and increase its productivity in irrigated rice," *Agricultural Water Management*, vol. 49, no. 1, pp. 11-30.

- [11] Shiqi Peng, 2011. "Water resources strategy and agricultural development in China," *Journal of Experimental Botany*, vol. 62, pp. 1709-1713, [Abstract/Free Full Text]
- [12] Li. B. and Peng. S., 2009. "China's agricultural water use report (1998-2007)," Beijing, China. [Abstract/Free Full Text]
- [13] A. Idrissa, 2008. "Effet of water and fertilizer on the growth and yield of summer maize under mulching condition," Ph.D. Thesis Hohai University, Nanjing, p. 140.
- [14] China Meteorological Administration, "Weather Information for Nanjing, Climatological Information." <http://www.worldweather.org/001/c00350f.htm#wxfore>.
- [15] Guo Xiang-ping, Yuan Jing, G. Feng, and C. Zhi-ping, 2006. "Preliminary research on water-catching and controllable irrigation technology of rice," vol. 4, pp. 537-541
- [16] C. P. Jahn, G. C., B. Khiev and S. Pheng, 1997. "Pest management in rice," in *In "Rice Production in Cambodia*, H. J. Nesbitt, Ed. Manila Philippines: International Rice Research Institute, pp. 83-91.
- [17] R. J. Hnatiuk, 1990. *Census of Australian vascular plants*, Series 11. Australia: Australian Flora and Fauna, p. 98. [Abstract/Free Full Text]
- [18] B. Sunderland, K.D de Snoo G.R. Dinter A Hance, T Helenius, J., Jepson, P., Kromp, B., Samu, F., Sotherton, N.W., Toft, S. and Ulber, 1995. "Density estimation for invertebrate predators in agroecosystems," *Acta Jutlandica*, vol. 70, pp. 133-164
- [19] APHA, 1995. (American Public Health Association), "Standard Methods for the Examination of Water and Wastewater," Washington DC.
- [20] Barrett Purcell & Associates, 1999. "Determining a framework, terms and definitions for water use efficiency in irrigation.," Australia. [Abstract/Free Full Text]
- [21] F. A. O. 2005. United Nations, "Formation sur l'entomologie appliquée aux méthodes de lutte contre les insectes nuisibles aux cultures et sur l'utilisation des pesticides," Paris, France. 10p. [Abstract/Free Full Text]
- [22] Mao Zhi and Cui Y.L., 2001. "Irrigation techniques of water-efficient and sustainable humper yield for paddy rice," Wuhan (China), 32p. [Abstract/Free Full Text]
- [23] B. Zwart, 2004. "Reviewed of measured crop productivity values for irrigated wheat, rice, cotton and maize," *Agric. Water Management*, vol. 69, pp. 115-133
- [24] VARTAPETIAN, B. B., AND M. B. JACKSON., 1997. "Plant adaptations to anaerobic stress." *Annals of Botany*, vol. 79, pp. 3-20,
- [25] M. Arshad and W. T. J. Frankenberger., 1990. "Production and stability of ethylene in soil," *Biology and Fertility of Soils*, vol. 10, pp. 29-34.,
- [26] C. Weining, 1993. "Study of the impacts on soil character by percolation from rice fields," Wuhan University of Hydraulic Engineering, China. [Abstract/Free Full Text]
- [27] S. K. De Datta and R. W. Herdt, 1983. "Weed control technology in irrigated rice," in: *Proceedings of the Conference on Weed Control in Rice*, pp. 89-108. [Abstract/Free Full Text]
- [28] P. Francis, 1980. "Mid-term Technical Report of Agronomist," Gazipur, 21p. [Abstract/Free Full Text]
- [29] S. K. De Datta and M. Z. Haque, 1982. "Weeds, weed problems and weed control in deepwater rice areas," pp. 427-442. [Abstract/Free Full Text]
- [30] Scott A.T., 1998. "Agricultural water management in the Mississippi Delta region of Arkansas," *Arkansas Agriculture Experiment Station. Research Bulletin*, vol. 959, pp. 14-62.
- [31] CALU Factsheet, 2009. "MAJOR PESTS, DISEASES AND WEEDS OF CEREAL CROPS," Buddsoddi, GERMANY, p. 21 [Abstract/Free Full Text]
- [32] Z. Zhang, B. Soulemane, X. Guo, C. X. Zhu, and Hui-lian, 2011. "Grain yield and weed-mediated pest incidence in paddy rice crops under controllable irrigation regimes," *International*

Journal of Food, Agriculture & Environment, vol. 9,
no. 3 & 4, pp. 100-110

- [33] R. R. John Wiley, 1993. *The ecological background of food production*, UK London: Ciba Foundation 77. Chichester (UK), p. 285. [Abstract/Free Full Text]

- [34] I. Amuwitagama, 2002. "Analysis of pest management methods used for Rice stem borer (*Scirpophaga incertulas*) in Sri Lanka based on the concept of Sustainable Development," Lund University, Sri Lanka, 44p.

Performance Analysis of Broadcasting Methods in Geocast Region for Vehicular Ad Hoc Networks

Sanjoy Das, D.K. Lobiyal

Abstract— In this paper, performance analysis of different broadcasting methods i.e. flooding or blind broadcast and probabilistic broadcasting inside the geocast region has been done. Our objective is to provide a comparative analysis between flooding and probabilistic methods with varying number of nodes in Vehicular Ad hoc Networks. Simulations have been conducted using the NS-2 simulator. For result analysis, we have used awk programming and Matlab. Different values of probability for probabilistic broadcast method have been considered to investigate an appropriate value that may give best results. The results show that probabilistic broadcasting method achieves maximum packet delivery ratio is 83.5 % when number of node is 196. In, sparsely populated network the packet delivery ratio for all cases is low. The minimum value of PDR obtained for sparsely populated network is 14.95%. From the result, it has also been observed that for better delivery ratio, message broadcasting should be done with minimum value of p for both the sparse and dense network.

Index Terms— Flooding, Probabilistic Broadcast, Ad hoc Networks, Vehicular Ad Hoc Network, Mobility Model, Packet Delivery Ratio.

1 INTRODUCTION

VANET is a special class of Mobile Ad hoc Network (MANET), where every node is a vehicle moving on the road.

In this network a node behaves like a router to relay a message from one node to another. In VANET mobility of vehicles depends on the structure of the geographical areas. VANET uses two types of communication methods- One from vehicle to vehicle (V2V) and the other is vehicle to fixed road side infrastructure (V2R). In both the methods vehicles can communicate to other vehicles or road side unit either directly or through multiple hops. It depends on the position of the vehicles. Further, the road side units (RSU) can also communicate with other RSU via single or multi hop. The RSU supports numerous applications like road safety, message delivery; maintaining connectivity by sending, receiving or forwarding data in the network. The main focus of the VANET is to provide real-time and safety applications for drivers and passengers. There are various types of safety features and services supported by VANET that are needed to be timely disseminated to a driver. Some of the applications are collision warnings, road sign alarms, blind turn warning, congested road notification, free flow tolling, parking availability notifi-

[1, 2].

By delivering these messages on time can minimize road accidents and save total journey time. The RSU can improve traffic management system by providing drivers and passengers with the above vital information. It is desirable that protocols should maintain the low end-to-end delay and, high delivery ratio, low overheads and minimum numbers of hops.

The rest of paper is organized as follows. Section 2 presents work related to the geocast protocols. In section 3 proposed model and an overview of flooding or blind broadcast and probabilistic broadcasting techniques is presented. In section 4 simulation environment and result analysis is discussed. Finally, the work presented in this paper is concluded in section 5.

2 RELATED WORK

Extensive works have been carried out by researchers, academicians and industries for successfully routing of messages in VANET. There are several research projects on VANET being carried out by researchers. Some of them are [CarTalk, Fleet-Net-Internet on the Road, NoW (Network on Wheel)] [1, 2] with the emphasis on deployment in the real world. The main focus of all these projects is to provide safety, comfort and timely dissemination of message from one location to another location. Some of the message delivery protocols proposed for VANET attempt to deliver a message to a geographic region rather than to a node. These protocols are called geocast routing. LAR [4], LBM [5], and GeoTORA [6] is modified TORA, GRID protocol is modified to GeoGRID [7], DREAM [8], GRUV [9], are few geocasting protocols. In [5] authors use flooding method but it limits the flooding to a small region called forwarding zone instead of whole area. The forwarding is computed based on the position of sender and geocast re-

-
- Sanjoy Das, Ph.D Scholar, School of Computer and Systems Sciences, Jawaharlal Nehru University, India, E-mail: sdas.jnu@gmail.com
 - D.K Lobiyal, Associate Professor, School of Computer and Systems Sciences, Jawaharlal Nehru University, India, E-mail: lobiyal@gmail.com

-cation, parking spot locator, internet connections facility, electronic toll collection, and a variety of multimedia services etc

gion. In [6] authors have improved the method proposed in [5] and incorporate it with TORA. Through simulation study, they have shown that this method reduces the overhead of geocast delivery, and maintain high accuracy in data delivery. All these protocols use simple flooding technique inside the geocast region for message delivery. The flooding technique is the simplest broadcasting method to deliver message inside a geographical region i.e. geocast region. Further, in simple flooding technique [3, 17], any vehicle receive a broadcast message for the first time, has the responsible to rebroadcast the message. In this method, number of transmissions increases with increasing number of nodes in the network. In [13] authors show a wide analysis of their proposed protocol Geographic Source Routing (GSR) with DSR, AODV for VANET in city scenarios. They have done simulation analysis of these protocols on realistic vehicular traffic for a particular city. The real city map is considered and converted to graph for the analysis. Their result shows that GSR performs better than DSR and AODV in terms of end-to-end delivery and latency. In [11] and [14] the authors proposed different modified LAR algorithms. They have modified the request zone. Through simulation, the authors have established that their proposed algorithms reduces route request overhead as compared to original LAR. The performance analysis shows that their method outperforms original LAR especially, in a dense and highly dynamic ad hoc network. In [12] the authors have proposed a greedy version of LAR protocol known as GLAR (Greedy Location-Aided Routing Protocol). This scheme improved the performance of LAR. In GLAR method, to find a route between source and destination, a baseline is drawn between them. The route request packets are broadcast within the request zone. The neighbouring node which has shortest distance towards baseline is selected as next broadcasting node. The authors considered various network performance parameters to compare LAR with GLAR. Their results revealed that GLAR reduces the number of route discovery packets and increases the average network route lifetime. In [10] authors have only considered the energy consumption parameter for performance analysis of LAR1 protocol with DSR and AODV in highly dense ad hoc networks. The results reported show that LAR1 perform better than DSR and AODV protocol in highly dense network. But in low density DSR performs better than others in term of energy consumption. In [18] the authors analysed the performance of LAR1 protocol in city scenario. Through extensive simulation they have shown the end-to-end delay is high in sparsely populated network but in densely populated network end-to-end delay is low. Most of these protocols use random waypoint mobility model for performance analysis. None of above protocol considered the grid structure for node deployment.

3 PROPOSED MODEL

We have considered the multi-hop environment, because it's very rare that source and destination node fall in each other transmission range. As there is no direct connectivity between source and destination node, to route the message interme-

mediate nodes plays a vital role. The intermediate nodes are act as relay node. We have considered highway scenario to deliver message from source to geocast region shown in fig-1. To deliver data to all the nodes inside the geocast region we have considered flooding and probabilistic techniques.

As demonstrated in this document, the numbering for sections upper case Arabic numerals, then upper case Arabic numerals, separated by periods. Initial paragraphs after the section title are not indented. Only the initial, introductory paragraph has a drop cap.

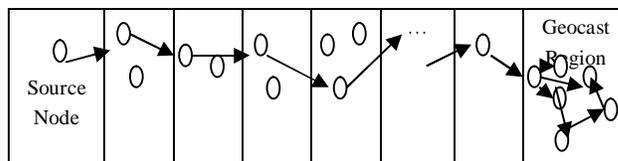


Fig 1. Simple Scenario of Geocast on Highway

3.1 Overview of Flooding and Probabilistic Techniques

The flooding or blind broadcast [17] is the simplest method to deliver a message to all nodes present in a specific area. This is the most guaranteed geocasting mechanism. Here, a node that receives a message for the first time; will retransmit it to all its neighbours. This method only guarantees that, a message will be definitely delivered to the destination in a connected network. Here, the packet delivery ratio is high, but the overhead is also very high. Suppose, n numbers of node are participated in message dissemination in the geocast region. As the number of nodes n increases no of packet to transmitted increases. It causes redundant data transmission and inefficient use of network resources. This method ensure that, node present in the geocast region receives a copy of a geocast packet. Sometimes it leads to broadcast storm problem [16] due to high contention, collisions and redundant rebroadcast of messages. To mitigate the broadcast storm problem some solutions related to VANET is proposed in [15].

3.1.1 Flooding Algorithm [17,19,20]

```

source node send packet (pkt) received at node  $n_i$ 
if  $n_i$  received P for first time
{
  if (neighbour  $n_i \neq \text{NULL}$ )
  broadcast (pkt) to its neighbour node
}
else
discard packet (pkt).
End.
    
```

The Probabilistic technique of broadcast [17] is a type of restricted flooding. To mitigate the shortcoming of flooding this method was introduced. In this method, upon receiving a non

duplicate packet nodes further rebroadcast with probability p .
where $(0 < p \leq 1)$.

3.1.2 Probabilistic Broadcast Algorithm [17,19,20]

```

Source node send packet (pkt) received at node  $n_i$ 
if  $n_i$  received pkt for first time
{
  if (neighbour  $n_i \neq \text{NULL}$ )
  choose value probability of P //  $0 \leq p \leq 1$ 
  broadcast (pkt) to its neighbour node with P
}
End.
```

4 SIMULATION ENVIRONMENTS AND RESULT ANALYSIS

The simulation has been carried out to evaluate the performances of simple flooding or blind broadcast and probabilistic broadcast protocols for VANETs by using the network simulator NS-2 [21]. The table 1 shows different simulation parameters and table 2 shows the different parameters values considered for simulation. The results for probabilistic broadcast have been presented in Table 3(a) and Table 3(b). In the results we have computed the packet delivery ratio for both the protocols. We have uses the awk programming and Matlab [22] for analyzing the simulation results. According to Fig.1 the geocast region we have considered is 500 m x 500 m. All the results presented are obtained as an average of 10 different simulation runs.

We have deployed the nodes in the simulation area on the basis of GRID structure. Where, the node placement starts at (0, 0). In this method each node is one GRID-UNIT away from its neighbors node. It is essential that the number of nodes must be square of an integer to support the GRID structure. The Grid Unit can be computed as:

$$\text{Grid Unit} = \frac{\text{val}(X)}{\sqrt{(\text{no of nodes})}} \tag{1}$$

TABLE 1

SIMULATION PARAMETERS

Parameter	Specifications
MAC Protocol	IEEE 802.11 DCF
Radio Propagation Model	Two-ray ground reflection model
Channel type	Wireless channel
Antenna model	Omni-directional

TABLE 2

VALUES OF SIMULATION PARAMETERS

Parameter	Values
Simulation Time	1000s

Simulation Area (X *Y)	500 m x500 m
Transmission Range	250 m
Bandwidth	2 Mbps

4.1 Packet Delivery Ratio

Packet delivery ratio is a very important metric to measure the performance of routing protocol. The performance of a protocol depends on various parameters chosen for the simulation. The major factors are packet size, no of nodes, transmission range and the structure of the network. The packet delivery ratio can be obtained from the total number of data packets arrived at destinations divided by the total data packets sent from sources.

$$\text{Packet Delivery Ratio} = \frac{\sum (\text{Total Packets send by all Source node})}{\sum (\text{Total packet send by all source node})}$$

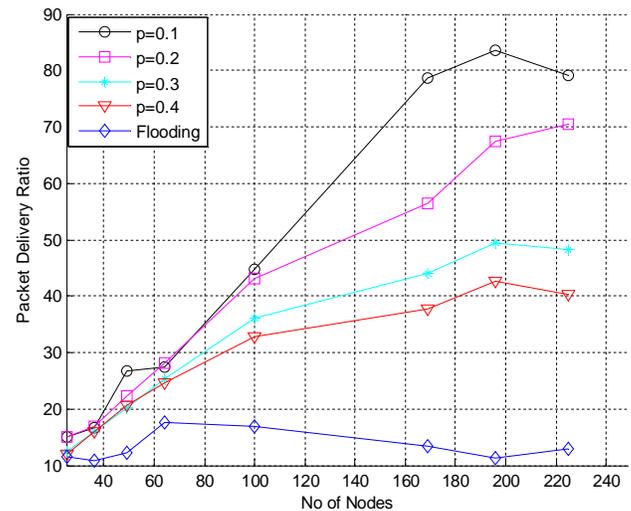


Fig 2. Packet Delivery Ratio of flooding and Probabilistic broadcast with $p=0.1, 0.2, 0.3, 0.4$.

Fig.2 shows the packet delivery ratio of flooding and probabilistic broadcasting techniques. We have shown PDR for flooding and probabilistic broadcasting with $p=.1, 0.2, 0.3, 0.4$. In the flooding method maximum value of PDR is 17.5231 when number of node is 64. For $p=0.1, 0.2, 0.3, 0.4$ packet delivery ratio gradually increases as the number of nodes increases. The maximum achievable PDR is 83.56629 when $p=0.1$ and number of node is 196.

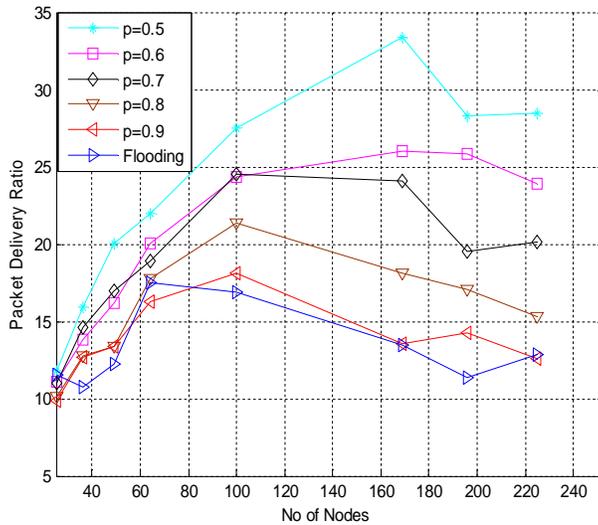


Fig 3. Packet Delivery Ratio of flooding and Probabilistic broadcast with p=0.5, 0.6, 0.7, 0.8, 0.9.

Fig.3 shows the packet delivery ratio of flooding and probabilistic broadcasting techniques. We have shown PDR for flooding and probabilistic broadcasting with p=0.5, 0.6, 0.7, 0.8, 0.9. For p=0.5, 0.6, 0.7, 0.8, 0.9 packet delivery ratio gradually increases as the number of nodes increases. When the number of nodes increases from 169 to onwards the PDR value starts decreasing for all values of p. The maximum achievable PDR value is 33.36514 when p=0.5 and number of node is 169.

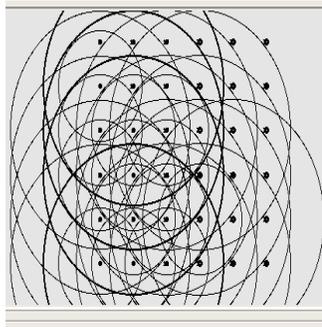


Fig 4. Shows the snapshot of simulation when n=36 and p=0.1.

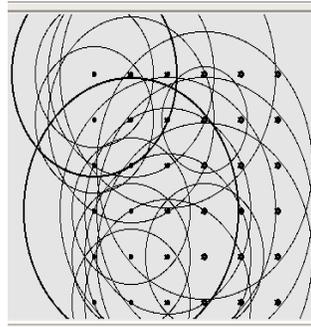


Fig 5. Shows the snapshot of flooding for n=36.

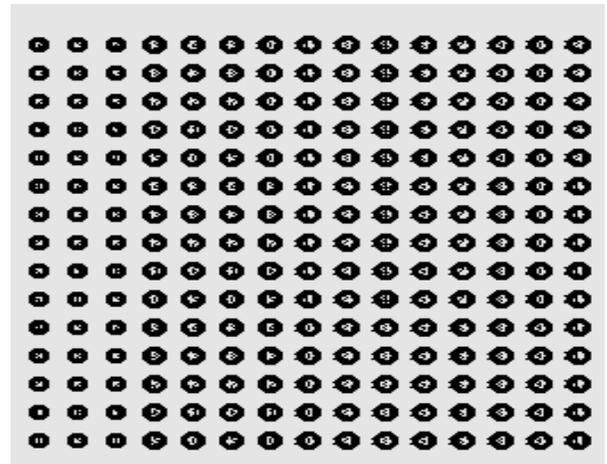


Fig 6. Snapshot of Node deployment for n=225.

In fig 4 shows the snapshot of simulation when n=36 and p=0.1 and fig 4 shows the snapshot of flooding algorithm for when number of node is 36. In fig 5 shows the how we have deployed the nodes in the simulation area when number of nodes is 225 are deployed.

TABLE 3(A)

VALUE OF PDR WITH DIFFERENT VALUE OF PROBABILITY AND NUMBER OF NODES

Broad-cast Probability	No of nodes			
	25	36	49	64
0.1	14.9583	16.58472	26.72001	27.37938
0.2	14.95237	16.79714	22.19474	28.23787
0.3	12.53139	15.98569	20.25287	25.41184
0.4	11.95905	16.02106	20.65038	24.54625
0.5	11.80366	15.96245	20.02624	21.96514
0.6	11.08429	13.80751	16.18794	20.00724
0.7	11.03823	14.6289	16.98467	18.87625
0.8	10.12938	12.73513	13.41681	17.72958
0.9	9.86544	12.67237	13.4129	16.26605

TABLE 3(B)

VALUE OF PDR WITH DIFFERENT VALUE OF PROBABILITY AND NUM-

BER OF NODES

Broad-cast Probabilistic	No of nodes			
	100	169	196	225
0.1	44.74569	78.58425	83.56629	79.10344
0.2	43.06935	56.50717	67.35748	70.43942
0.3	36.19645	43.92315	49.50677	48.17908
0.4	32.79958	37.65447	42.62175	40.40304
0.5	27.49097	33.36514	28.26565	28.46729
0.6	24.33346	26.00031	25.8327	23.87187
0.7	24.5071	24.1111	19.51238	20.11281
0.8	21.39716	18.15165	17.10252	15.27812
0.9	18.08374	13.55761	14.26087	12.5451

5 CONCLUSIONS

In this paper we have analyzed the performance of flooding and probabilistic broadcast protocols to deliver messages inside a geocast region for vehicular ad hoc networks. The performances of these protocols are analysed for varying node density and different value of p for probabilistic broadcast. From the result analysis it is clearly evident that when the network is sparsely populated, the successful delivery of message is nearly 15% for p=0.1. But for densely populated network the highest value of PDR is 83.5 for n=196 and p= 0.1. Further, we conclude that probabilistic broadcast protocol outperforms flooding in all the cases except p=0.9 and n=64. It is also observed that for better delivery ratio, message broadcasting should be done with minimum value of p.

References

- [1] Hassnaa Moustafa and Yan Zhang, "Vehicular networks: Techniques, Standards, and Applications", CRC Press, Taylor & Francis Group, Boca Raton London, New York, 2009.
- [2] Marc Emmemmann, Bernd Bochow and Christopher Kellum, "Vehicular Networking Automotive Applications and beyond" A John Wiley and Sons, Ltd, Publication, United Kingdom, 2010.
- [3] Y. B. Ko and N. Vaidya, "Flooding-Based Geocasting Protocols for Mobile Ad Hoc Networks", Journal of MOBILE NETWORKS AND APPLICATIONS, VOL 7; NUMBER 6, pp. 471-480,2002.
- [4] Y. B. Ko and N. Vaidya, "Location-aided routing (LAR) in mobile ad hoc networks," in Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM'98), pp. 66-75, 1998.
- [5] Y. B. Ko and N. Vaidya, "Geocasting in Mobile Ad-hoc Networks: Location-Based Multicast Algorithms", In 2nd IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, Louisiana, pp. 101-110, February 1999.
- [6] Y. B. Ko and N. Vaidya, "GeoTORA: A Protocol for Geocasting in Mobile Ad Hoc Networks", In IEEE International Conference on Network Protocols, Osaka, Japan, pp.240-250,2000.
- [7] W.H Liao, Y. C Tseng, K. L Lo and J. P Sheu, "GeoGRID: A Geocasting Protocol for Mobile Ad Hoc Networks Based on GRID", Journal of Internet Tech. Vol.1 No.2, pp. 23-32, 2000.
- [8] Stefano Basagni, Imrich Chlamtac, Violet R. Syrotiuk, and Barry A. Woodward, "A distance routing effect algorithm for mobility (DREAM)", In MobiCom'98:Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking, pp. 76-84, 1998.
- [9] Guoqing Zhang,, Wu. Chen, Zhong Xu,, Hong Liang, and Li Gao Dejun Mu, "Geocast Routing in Urban Vehicular Ad Hoc Networks", in R. Lee, and G. Hu, H. Miao, Eds. Computer and Information Science 2009, SCI 208, Springer-Verlag Berlin Heidelberg,pp.23-31, 2009.
- [10] E.Ahvar and M. Fathy, "Performance evaluation of routing protocols for high density ad hoc networks based on energy consumption by GlomoSim simulator" in the Proceedings Of World Academy Of Science, Engineering And Technology, Volume 23, pp. 97-100, August 2007.
- [11] Senouci, S.M. and Rasheed, T.M., in IFIP International Federation for Information Processing, Volume 229, Network Control and Engineering for QoS, Security, and Mobility, IV, ed. Gaiti, D., (Boston: Springer), pp. 137-146, 2007.
- [12] Neng-Chung Wang, Jong-Shin Chen, Yung-Fa Huang, and Si-Ming Wang, "A Greedy Location Aided Routing Protocol for Mobile Ad Hoc Networks". In proceedings of the 8th WSEAS International Conference on Applied Computer and Applied Computational Science (ACACOS '09) Hangzhou, China, pp 175-180, 2009.
- [13] Christian Lochert, Hannes Hartenstein Jing Tian, Holger Fùßler, Dagmar Hermann, and Martin Mauve, "A Routing Strategy for Vehicular Ad Hoc Networks in City Environments", Proceedings of IEEE Intelligent Vehicles Symposium, pp.156-161,2003.
- [14] F.De Rango, A.ha, A.Molinaro,and S. Marano, "A Modified Location Aided Routing Protocol for the Reduction of Control Overhead in Ad-hoc Wireless Networks", ICT2003, vol.2, February 23 -March 1, pp. 1033-1037, 2003.
- [15] N. Wisitpongphan, O. K. Tonguz, J. S. Parikh, P. Mudalige, F. Bai, and V. Sadekar , "Broadcast storm mitigation techniques in vehicular ad hoc networks", IEEE Wireless Communications. Vol. 14, pp. 84-94, 2007.
- [16] Sze-Yao Ni, Yu-Chee Tseng, Yuh-Shyan Chen, and Jang-Ping Sheu The Broadcast Storm Problem in a Mobile Ad Hoc Network, in Proceedings of ACM MOBICOM ' 99 , Seattle, WA , pp. 151 - 162, 1999 .
- [17] Jun Zheng and Abbas Jamalipour, "WIRELESS SENSOR NETWORKS A Networking Perspective", John Wiley & Sons, Inc., Hoboken, New Jersey, pp-166, 2009.
- [18] S. Das and D.K. Lobiyal, "A Performance Analysis of LAR Protocol for Vehicular Ad Hoc Networks in City Scenarios", Proc. of Int. Conf. on Advances in Computer Engineering, pp.162-166. DOI: 02.ACE.2011.02, 2011.
- [19] Y. Sasson, D. Cavin, A. Schiper, "Probabilistic broadcast for flooding in wireless mobile ad hoc networks", Proc. IEEE Wireless Communications & Networking Conference (WCNC 2003), pp.1124-1130, 2003.

- [20] Y. Sasson, D. Cavin, A. Schiper, "Probabilistic broadcast for flooding in wireless mobile ad hoc networks", EPFL Technical Report IC/2002/54, Swiss Federal Institute of Technology (EPFL), 2002.
- [21] The Network Simulator. <http://www.isi.edu/nsnam/ns/>.
- [22] The Math Works: <http://www.mathworks.com>

The Influence of Roughness on the Wear and Friction Coefficient under dry and lubricated sliding

Riyadh A. Al-Samarai¹, Haftirman¹, Khiarel Rafezi Ahmad², Y. Al-Douri³

Abstract— The aim of the present research was to the effect of surface roughness of aluminum-silicon casting alloy on the friction and wear is investigated. Various surface roughness average (Ra) of different degrees were verified via different grades of grinding, polishing and have been evaluated using a pin-on-disc as well as three different loads 10, 20 and 30 N, speeds 200, 300 and 400 rpm and relative humidity 70%. Different surface preparation techniques are resulted in different surface roughness of (Ra) = (4,6) μm . The monitor effects on the friction coefficient and wear are through the test dry and lubricated sliding. It was noted that the weighted and volumetric wear rate decreases as degree of roughness decreases, as well as coefficient of friction is considered as a function of the stability state. This paper attempts to bridge the gap between the damage mode, sliding conditions and surface roughness to provide an approach to evaluate the surface finishing as a factor in friction and wear damage processes.

Index Terms — Surface roughness, Lubrication, Dry sliding, Casting alloy, Coefficient of friction, Wear rate, Damage.

1. INTRODUCTION

Surface roughness is a major problem produced during the production process and greatly affect the quality of the product [1]. Can affect the surface roughness during the process of sliding through the contact between the surfaces at the Tops and this leads to cut and deformation these peaks during the slide and thus can cause economic damage so it is best to be a better understanding of the roughness[2] as well as affect the surface roughness coefficient of friction on the side by side with the frequencies and vibration incurred during sliding between the surfaces, and there are variables affecting the coefficient of friction when sliding[3][4]. is important to know the quality of the roughness of surfaces that require a process of coating and treatment to protect it from damage[5] That the installation and texture surface can increase the load hydrodynamic as well as increasing the thickness of film when the structure of surfaces of linear and moderate and thus increases the life of surfaces and reduce the cost [6]. of either rough surfaces usually lead to speed the process of wear and damage surfaces when working must take into account the mechanical properties and factors affecting when sliding between surfaces[7]. In these experiments we will notice the impact of this roughness is actually a coefficient of friction and wear rate between the surfaces sliding at the contact between the areas of nominal and actual handling of this effect by carrying out treatments

the surface is the manufacturing process of the basic surfaces and the surface termination phase[8]. Menezes et al. [9]. have been studied the effects of roughness parameters on the friction of aluminum alloy under conditions of lubrication. They concluded that the coefficient of friction and wear are depending on the roughness It is important to know the quality of the surfaces roughness that requires a process of coating and treatment to protect it of damage [10]. Karpenko and Akay [11] have been studied the effect of roughness between two surfaces using an algorithm to calculate the coefficient of friction between them. They concluded that there is a flexible deformation and shearing resistance depend on external loads, mechanical properties and topography surfaces to give the approximate limits of influence. To increase the process of wear and damaged surfaces, we must take into consideration the mechanical properties affect the process [12]. . Chowdhury and Maksud [13] have been searched the effect of humidity on surface roughness and found that the friction is very high at low roughness and tends to be increased at high horizontal vibration. Wieleba [14] had been studied roughness and stiffness of composite materials against steel and showed its effect on the friction and wear. Al-Si alloys have been studied interestingly, but did not take into consideration the roughness study. Xing et al. [15] have been prepared the hypereutectic Al-17.5Si (wt pct) and Al-25Si (wt pct) alloys with various content of rare earth Er by conventional casting technique. They investigated the effect of Er on the microstructure and properties of hypereutectic Al-Si alloys using optical microscopy, scanning electron microscopy (SEM) as well as friction and wear tests and noticed an improvement of the anti-wear properties and the friction coefficient of the hypereutectic Al-Si alloys. Finally, Li et al. [16] have been prepared and fabricated hypereutectic Al-Si alloys by hot extrusion. They investigated the tensile fracture me-

- ¹School of Mechatronic Engineering, University Malaysia Perlis, Ulu Pau, Perlis, Malaysia
- ²School of Materials Engineering, University Malaysia Perlis, 02600 Arau, Perlis, Malaysia
- ³Institute of Nano Electronic Engineering, University Malaysia Perlis, 01000 Kangar, Perlis, Malaysia

on the surface of the coating, and other processes that help in smoothing the surface and the things that affect the advance on

chanisms using SEM and proved an improvement of Al-Si alloy and wear resistance due to silicon particles refining.

In this work, to best of our knowledge, there is not available in the literature investigation of roughness on the wear rate with sliding distance and velocity of Al-Si casting alloy, on the volumetric wear rate and friction coefficient with normal load. It was attempted to bridge the gap between the damages resulting from the sliding surfaces and surface roughness effect on the friction and wear to provide an approach for evaluation the surface finishing operations. All of these are divided into the followings: Section 2 displays the experimental process, while results and discussion are given in section 3. Finally, section 4 concluded the obtained results.

2. Experimental Procedure

For the purpose of this investigation, aimed at investigating influence of surface preparation on roughness parameters and correlation between roughness parameters and friction and wear, A pin-on-disk tribological test rig was used for the investigation shown in Fig. 1 The upper specimen was a fixed (10 mm) diameter Al-16Si casting alloy $R_a = (4,6) \pm 0.03 \mu\text{m}$, $H_v = 112.65 \pm 12 \text{kg/mm}^2$, on disc, made of AISI 1045 steel ($R_a = 0.15 \pm 0.05 \mu\text{m}$, $H_v = 312 \pm 20 \text{kg/mm}^2$), were tested. The applied load was (10,20,30) N and the sliding speed, (200,300,400) rpm (1.32,1.885.3) m/s. before the start of effective wear samples of aluminum silicon disc was cleaned and dried using cotton and acetone as the weight of the samples was measured using a digital balance and recording the values before and after the test in each test is calculated sliding distance. The wear tracks were observed by scanning electron microscopy (SEM) combined with energy dispersive X-ray spectroscopy (EDX). Surface roughness was also measured by a stylus surface analyzer, with the effective measure length 0.350 mm and the cutoff length, 0.05mm. The work hardening of contact surfaces due to the friction shear was identified by their hardness increase as determined by micro-Vickers indentation test.



Fig. 1 Pin-on-Disc wear testing machine

Chemical analysis was conducted for the aluminum-silicon casting alloy, also density, hardness and tensile strength are studied due to its widely used in industry, particularly in pistons as well as the cylinders, the resulted chemical analysis is given in the Table 1, and testing of mechanical properties are cleared in Table 2. while the surface roughness parameters (R_a) are calculated in Table 3.

Table 1: Compositional analysis of Al-Si casting alloy.

Si	Mg	Cu	Fe	Ni	Mn	Sn	Pb	Zn
16.69	1.176	1.30	1.13	1.22	0.02	0.012	0.026	0.01

Table 2: Investigated hardness, density and tensile strength of Al-Si casting alloy.

Hardness	112.65 VHN
Density	2.72 gm/cc
Tensile strength	250 MPa

Table 3: The calculated surface roughness average (R_a) of Al-Si casting alloy.

Reference surface Sample area	R_a (μm)
Sample 1	4
Sample 2	6

The microscopic structure and the composition of microscopic samples are examined and shown in Fig. 2. It is shown that Si particles are distributed uniformly, while the Si seems a bulk, we have used the following materials manifesting:

- 190 ml of water distil
- 3 ml from hydrochloric
- 2 ml from hydrofluoric acid

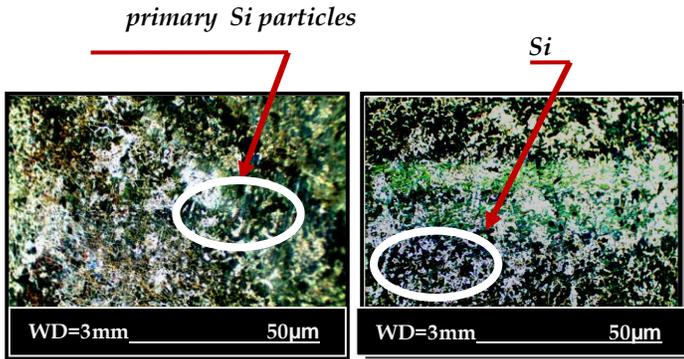


Fig. 2 The SEM images of microscopic structures of Al-Si casting alloy.

Al-Si casting alloy. Was cleaned then starting an actual test of wear process and the hard substance acetone with cotton and dry well was registered height and weight of samples accurately to provide a very precise and digital recording of all data using a stopwatch to set a time slip and post-test .

Wear rate was estimated by measuring the mass loss in the specimen after each test and mass loss, (ΔW) in the specimen was obtained. Cares have been taken after each test to avoid entrapment of wear debris in the specimen. It is calculated to the mass loss to sliding distance ($S.D$) using:

$$W.R = \Delta W / S.D \quad (1)$$

The volumetric wear rate Wv of the composite is related to density (ρ) and the abrading time (t), using:

$$Wv = \Delta W / \rho t \quad (2)$$

The friction force was measured for each pass and then averaged over the total number of passes for each wear test. The average value of coefficient of friction, μ of composite was calculated from

$$\mu = Ff / Fn \quad (3)$$

where Ff is the average friction force and Fn is the applied load with an assumption that the temperature is constant at 31 °C.

$$Ws = Wv / S.S \quad (4)$$

where S.S is the sliding velocity.

3. Results and discussion

The results of tests conducted on different surface roughness of (Ra) = (4,6) μm under dry sliding speed and conditions of (1.32, 1.885, 3) m/sec within the range of different

loads; (10, 20 and 30) N, and note in Figure 3.(a) that the value of surface roughness (Ra) = 6 μm gave highest percentage in the rate of wear weighted $W.R = 0.1473 \times 10^{-6}$ (N/m), at the speed of low compared with the amount of roughness to the surface of the (Ra) = 4 μm , the other $Ra = 4 \mu m$ gives $W.R = 0.1262 \times 10^{-6}$ (N/m) and goes down with the passage of time sliding the amount of roughness (Ra) = 4 μm had given the rate of wear is less compared with the surface of the former reason is attributed to increase the surface roughness reduces the area of contact real are concentration load only in the areas of contact between the surfaces and gets broken layer oxide and cause an adhesion metal is strong and therefore, the force required to cut notches related to higher than the force required to cut the bonds of atomic alloy, This effect continued for all the loads used in the test In Figure 3. (b) When comparing these results between the dry test with the test Note the presence of lubrication to reduce all the values where we got on the rate of wear weighted $W.R = 0.1052 \times 10^{-6}$ (N/m), 0.0842 $\times 10^{-6}$ (N/m) respectively.

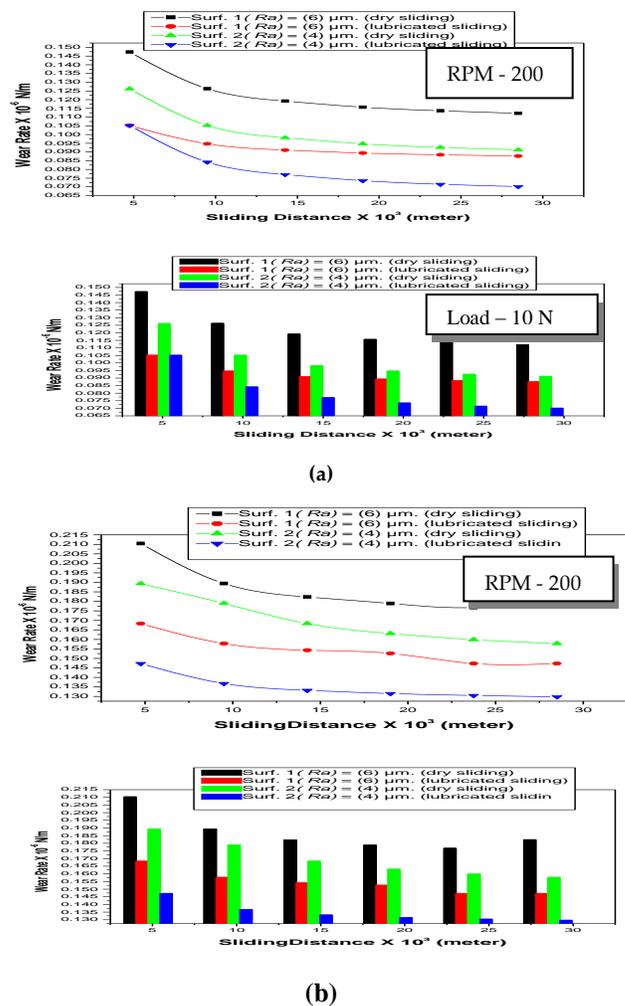
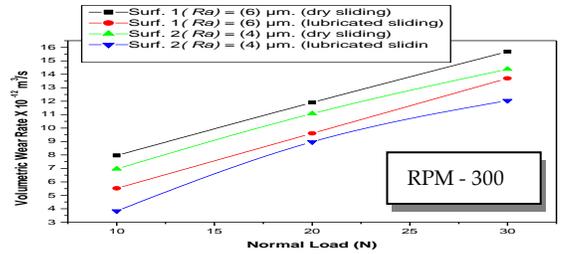
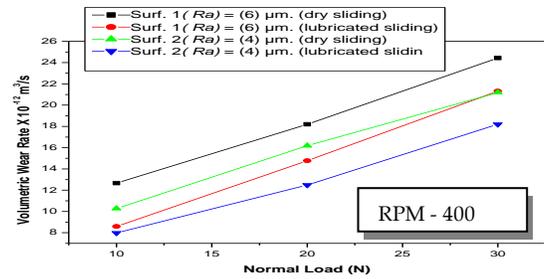


Fig. 3 Variation of surface roughness $Ra = 4$ and $6 \mu m$ and wear rate with sliding distance of 200 rpm at (a) 10 N, (b) 20 N. under dry and lubricated Conditions

In Figure 4. (a),(b) and (c) we note the low rate of volumetric wear with a low surface roughness, where it decreased from $Wv = 5.393 \times 10^{-12} \text{ (m}^3\text{/sec)}$ to $4.918 \times 10^{-12} \text{ (m}^3\text{/sec)}$ respectively and these values are also lower when compared with the test lubrication, where the presence of oil reached is due to the formation of a film led to lower the temperature and reduce the contact between the surfaces and thus reducing the friction, which reduced the loss of the metal weighted, as well as note that the amount of load significantly affect the specific of wear and this is clear, where the shed loads low during the slide will lead to the formation of film protector reduces the contact between the surfaces and this is the force required to cut the interdependence between the bumps less than the force required cut the ties of atomic alloy and therefore less wear specific is the qualitative increase of load showed the results of the surface at $Ra = 6 \mu\text{m}$ has amount of specific wear equals to $W_s = 3.410 \times 10^{-13} \text{ (m}^3\text{/N-m)}$. and fell to the amount slightly at high speed compared with the amount of roughness $(Ra) = 4 \mu\text{m}$, shows that the amount of loss of quality depends on the roughness of the surface as well as increases with roughness and less signal decreases, as well as can be seen that the reason for lack of wear resistance at the beginning of the test is to be a thin layer of material corroded between bumps the disk surface and settle over time and that the continuation of the process of sliding lead to emotional to the surface of the sample and the low rate of wear of the qualitative and by roughness of the surface.

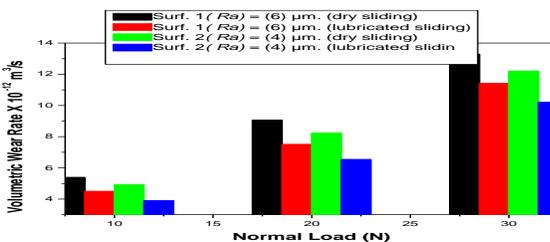
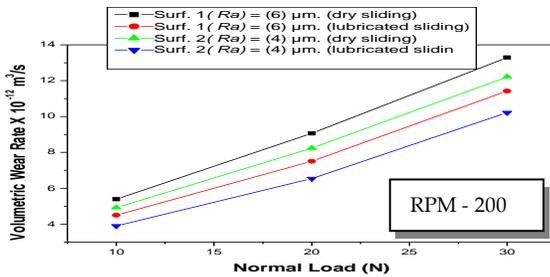


(b)



(c)

Fig. 4. Variation of surface roughness $R_a = 4$ and $6 \mu\text{m}$ and volumetric wear rate with normal load of 10,20 and 30 N at (a) 200 rpm, (b) 300 rpm and (c) 400 rpm. under dry and lubricated Conditions.



(a)

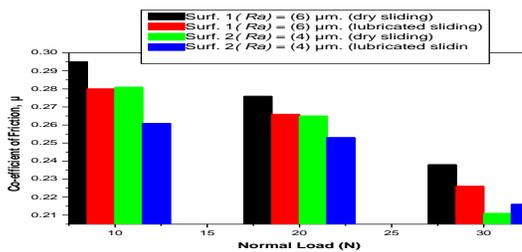
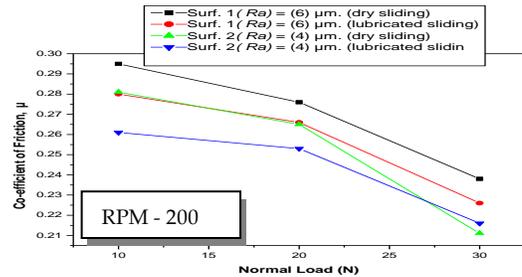


Fig. 5 Variation of surface roughness $R_a = 4$ and $6 \mu\text{m}$ and coefficient of friction with normal load of 200 rpm at (10) N. under dry and lubricated Conditions

Figure 5. that the surface roughness increases the friction coefficient and for all the loads that the coefficient of friction represents the percentage of friction force to the load hanging and that the increase in load lead to the flow of plastic the protrusions surfaces slippery and then an increase in the area of contact of real and this increases the area of connections and therefore we need to force cut is bigger than lead to increased friction force over time

increase the temperature of the surfaces and increase the flow of plastic to the protrusions, leading to flattening gradually protrusions become surfaces smooth and for the case become surfaces smooth and obtain steady-state and that higher speed leads to high temperature instant during the slide, leading to softer bumps, which reduces the shear force required to cut the connections and this leads to reduce the friction coefficient and when the test with a lubricant and all forms of the results were better and dropped all values for the wear rate and volumetric and weighted as well as specific addition to the coefficient of

friction and be because they are to the layer of film between the surfaces reduces the contact between them, leading to lower the temperature. while Low coefficient of friction with increasing load and sliding speed of the lubrication situation, compared with the dry state When test conditions similar. The micrographs wear surface increases with increasing load as shown in Fig. 6. It is observed for the same velocity, the wear rate on the surface is increased as load increases.

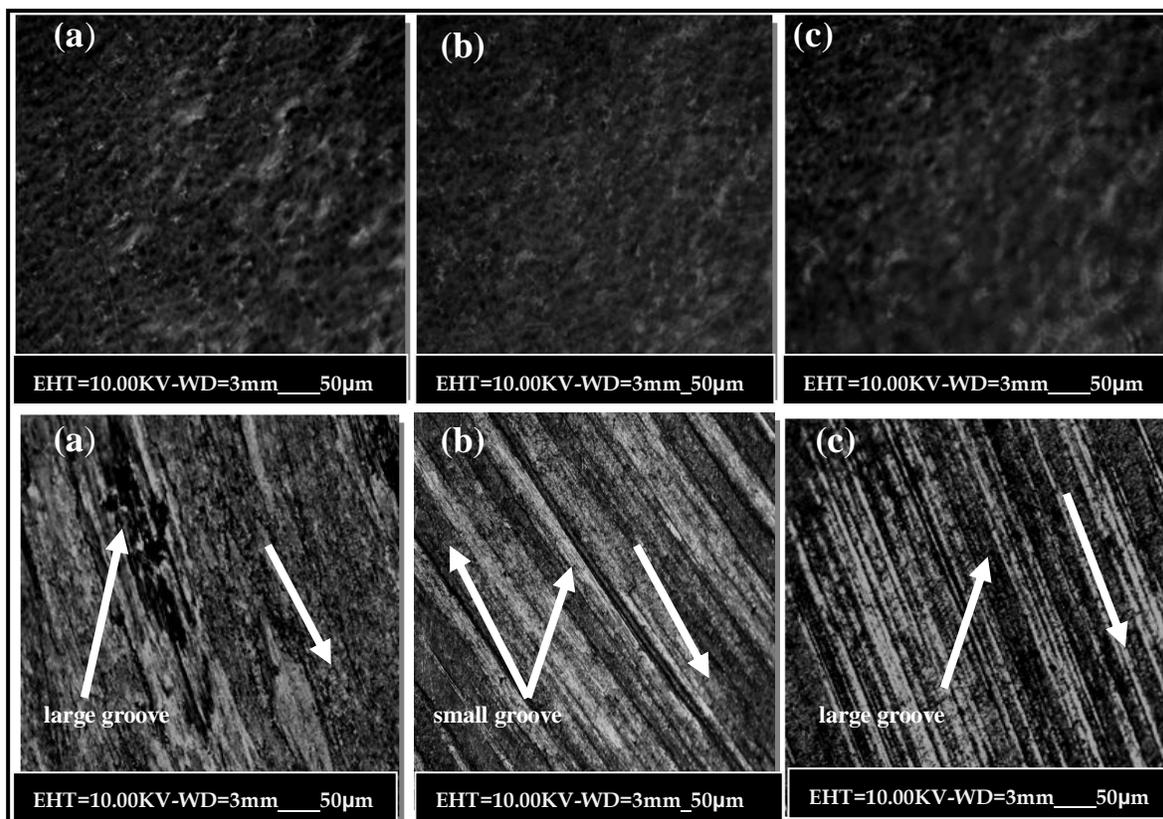


Fig. 6. The SEM images of micrographs before and after wear surface of 400 rpm and load (a) 10 N, (b) 20 N and (c) 30 N. under lubricated Conditions.

4 CONCLUSION

In summary, using the investigated results, we concluded the followings:

- The use of lubrication has led to lower wear rate the weighted and volumetric and specific, compared with dry sliding tests With low surface roughness.
- For dry contact, the friction conditions increase as the surface roughness increases. and low status of lubrication at the same test conditions.
- The roughness parameter of Al-Si casting alloy attributes to the shape of asperities of Ra and has the strongest influence on the wear rate.
- The wear rate increases as load and roughness average increase, while it correlates inversely with sliding distance.
- The specific wear rate decreases as load increases, it correlates inversely with sliding velocity, which is attributed to thin film of material eroded between bumps of the disk surface.
- The smoothness of surface as well as sliding distance reduces the volumetric wear rate due to lacking wear resistance of material thin film.

- The friction increases as the surface roughness increases, it correlates inversely with load.
- The higher value of friction is attributed to the higher value of R_a .

- forming technique, *J. Alloys & Comp.* 287, pp.284–294. 1999.
- [20] Men. PL, K. KSV. Surface texture and thus roughness parameters influence coefficient; 33:pp. 181–190. , 2008.
- [21] Ku KJ, Math TG. Interface roughness effect on friction map under fretting contact Conditions'. *Tribo. Intern*; 43:pp.1500- 1507. 2010.

ACKNOWLEDGMENT

I wish to thank everyone who contributed to the success of this work and the resident on it.

REFERENCES

- [1] Ro. t S. M. ski, Sc .D Bishop & Associates Inc ,Connector Degradation Mechanisms: Wear ,Bishop & Associates, Inc. 2010.
- [2] K.J. Kubiak ., T.G. Math., S. F., Interface rough. effect on friction map under fretting contact cond ., France *Tribo.*;43:pp. 1500- 507. 2010.
- [3] P.L. Menezes, Kishre and S. V.Influence of roughness parameters on coefficient of friction. Part 3, India;33:pp. 181–190. , 2008.
- [4] M. A. Chowdhury, M. Helali., *Mechanical Engineering, The Effect of Relative Humidity and Roug . on the Friction.* 2:pp.128-135. 2008.
- [5] S. Ekinci, A. Akdemir and M.T.Demirci3, university of Selcuk, Effect of Surface Roughness of Salt Bath Nitrided., *Techno* , Turkey2011.
- [6] R. Takat Dr. ,Effects of small-scale texturing on ring/liner friction Victor , Sacramento CA, Wong ICEF 2006.
- [7] K. Ka., M. M. Noor, M. M. Raman, *Mechanical Engineering, University Malaysia Pahang, European J. of Scientific Research*, 25: pp.250-256. , 2009.
- [8] A. Le Boot, E. Boo-Chakrat G. Micron, Dissipation of vibration in rough contact Laboratories., France , *Trib . Letters* .41:pp. 47-53. 2011.
- [9] W. R.B., I wabuchi A., 1985. The effect of ion implantation on the fretting. New York, 471–484.
- [10] Ka. K., Noor M., Raman M.M.,. Surface Roughness Prediction Model 6061-T6. *J. Scientific Research*. 25, pp.250-256. 2009.
- [11] Kar.Y.A., Akay A., *Mechanical Engineering, A numerical model of friction between rough surfaces.* *Tribo. Inter.* 1. 34, pp.531 - 545. 2001.
- [12] Tak. R . Effects of small-scale texturing on ring liner friction Victor, Sacramento CA, USA. 2006.
- [13] Cho. A.M., Helali M.,. The Effect of Relative Humidity and Roughness. *J. Mechanical Eng.* 2, pp.128-135,2008.
- [14] Wie . W., The statistical correlation of the coefficient of friction and wear rate of PTFE composites. *J. Wear*; 252, pp.719–729. 2002.
- [15] Xing P., Gao B., Zhuang Y., Liu K. , Tu G., On the medication of hypereutectic Al- Si alloys using earth Er, *Acta Metall. Sin. (Engl. Lett.)*. 23, pp.327-333. 2010.
- [16] Li Y.Y., Zhang D., Ngat T.L., Zhang W., Rapidly solidified hypereutectic Al-Si alloys prepared by powder hot extrusion, *Trans. Non ferrous Met. Soc. China.* 12, pp.878-881,2002.
- [17] Gupta M. Ling S. Microstructure and mechanical properties of hypo/hypereutectic Al-Si alloys synthesized using a near-net shape forming technique, *J. Alloys Com.*; 287:pp.284–294,1999.
- [18] Li YY. Zhang D. Ngat TL. Zhang W. Rapidly solidified hypereutectic Al-Si alloys prepared by powder hot extrusion, *Trans. Nonferrous Met. Soc. China*;12:pp.878-881. , 2002.
- [19] Gu. M., Ling S..Microstructure and mechanical properties of hypo/hypereutectic Al-Si alloys synthesized using a near-net shape

Critical Factors Facing Implementation of Power Quality Program Framework: Case Study-Libyan Electrical Distribution Networks

S. S. Sultan, M. K. Darwish

Abstract— Researches and concerns in power quality (PQ) gained significant momentum in the field of power electronics systems over the last two decades globally. This sudden increase in the number of concerns over power quality problems is a result of the huge increase in the use of non-linear loads. The purpose of this paper is to present a statistical analysis of power quality surveys conducted in Libyan Distribution Networks. It is also to explore the obstructions faced by Libyan distribution networks in implementing a power quality program (PQP) as well as to state the benefits, which would accrue by implementing a PQP, which would make a major impact on the distribution networks. In order to achieve these objectives, an extensive literature review was conducted to understand the barriers and benefits of implementing a PQP, followed by a power quality survey questionnaire and interviews. Data were collected from Libyan distribution networks (LDNs), both from departments and individual staff members. Both SPSS 15.1 and Nvivo 9 were used in performing the analysis. The results revealed that no power quality program exists. Out of 16 barriers, 12 were statistically significant different since the P value <0.05), which indicated that Libya distribution systems have already surmounted a few of the barriers to implementing a PQP effectively. The overall benefits of PQP implementation, which would have a positive impact on LDNs, are 11 benefits. The analysis also shows the level of awareness of power quality issues with the aim of generating a power quality framework which can be used as guidelines in the field of power quality. The findings of this paper are applied to build a PQP framework guideline to be implemented in LDNs. The PQP framework is consisted three essential phases. Phase one designed to increase the awareness level. Phase two is involved in preparation of PQP, which contains seven crucial requirements. Phase three is designed to prevent the outstanding problems from phase 1 and 2 of not reoccurring again to determine both the weaknesses and obstacles facing the implementation. The progress of this framework and moving through from phase to other will be depended on the level of awareness, knowledge, skills gained respectively after each phase performed.

Index Terms— Power Quality Surveys, Statistical Analysis, PQP Barriers, PQP Benefits, Libyan Distribution Networks, PQP Framework guideline.

1 INTRODUCTION

Since 1980, power quality issues have been causing disturbances to distribution systems internationally, and have attracted global concern [1], [2], [3], [4]. This rapid increase in power quality issues was due to deregulation, of electrical power industry [5], [6]. Furthermore, Electronic devices are one critical factor, which have brought significantly bad power quality variations to power systems supplies [7]. Demand for electricity is another factor causing poor power quality, due to the poor network design, which does not accommodate for economic growth [8]. Lack of awareness of the concept of power quality is one more aspect, which propagated power quality events. As a result, since the early 1990s, tackling these events has been a priority for utilities [9], [10]. Therefore, these issues considered as essential concerns for both the utilities and users [11], [12], [13]. Nowadays, these issues have driven both the electrical companies and end users to pay more attention for better understanding of power quality problems [14], [15]. Hence, some utilities describe power quality as critical to business, and have started to implement different programs to solve the problems [16]. The target is to determine the actual level of

power quality differences, and the main sources within distribution systems. On the other hand, some electrical

companies believe that it is not an area of concern. This is due to the fact they do not anticipate rapid growth of non-linear loads in the future [17]. However, it seems that there is a lack of understanding regarding the causes of these problems. Both electrical companies and end users' views are different [18]. Some electrical companies are forced to solve these problems due the huge number of complaints received from end users and also due to the large costs associated with poor power quality (for example: insurance claim) [19].

A large number of publications have been analyzed in this study to provide a comprehensive review of power quality surveys regarding the implementation of Power Quality Programs (PQPs) in the USA and Europe. However, recent studies [20], [21] shows that some countries like north of Africa and the Middle East are predicting to have a huge growth in power generation, transmission and distribution. This growth is not matched by similar growth in power quality awareness programs. This paper focuses mainly on the power quality survey in one of these fast growing North African countries (Libya).

The rapid growth of the Libyan economy began in 1999 [22]. Therefore, since the early 1999, tackling power quality events has been a priority for Libyan distribution networks (LDNs) [23]. Thus, the increase in peak load was not as rapid as it is nowadays; it was 5,964 MW in 2010, and expected to increase to 18,417 MW by 2025 [24]. Moreover, the level of power delivered was not at its worst level, this is mainly because sensitive equipment's were not yet introduced widely before 1999.

2 POWER QUALITY PROGRAMS (PQPs)

PQPs are particularly successful in developed countries rather than developing countries, due to the rapid adoption of sophisticated technology, as well as the higher level of PQ awareness among most of the end users, who recognize its importance. Furthermore, power suppliers in developed regions are trying to establish a high level of PQ standards in a short time, due to pressure from large industrial customers, as the use of sophisticated equipment increases [25].

In contrast, utilities in less developed countries are being pushed by the introduction of new technology from developed countries to improve and address their PQ issues. In response, some distribution companies have contracted a third party to solve PQ issues for their end users satisfaction; this is due to the inability of their engineers and technicians, who lack the skills and experience to solve these problems [17]. Therefore, government-controlled utilities are detached from the situation with regard to PQ issues. The failure to implement PQPs by some distribution utilities in developing countries have resulted in their supplying free power to their customers.

The distribution utilities in less developed countries are not worried about the quality of the power they provide to their clients. They believe that PQ has matured to the point, where it will not be of any importance in the future; moreover, their customers want only to be supplied with electricity, and are not concerned over quality [26]. Therefore, managers from distribution companies have concluded that some international electricity companies view implementing PQP as a business, rather than concerning themselves with issues of power distribution systems [17].

This study is the first to investigate the barriers and benefits of PQP within Libyan distribution systems. It contributes by providing an insight into the overall efforts needed to implement PQP framework and the main reasons underlying its failure. LDNs are among those systems facing poor power quality in under-developed countries. Unfortunately, statistical data show that in the last two decades, LDNs have not implemented power quality program [24]. This is mainly because there is no power quality department established yet, to influence the measurement of power quality disturbances (PQDs). This absence of a power quality department is due to lack of awareness on the part of top management regarding the importance of power quality. As a result, LDNs have faced very significant difficulties in implementing PQP. In addition,

lack of power quality awareness has led LDNs to face twelve significant difficulties through not implementing PQP [23], [27].

Four main factors of PQP barriers were determined from this study, namely; lack of awareness (lack of staff awareness, skills and experience, lack of end users' awareness, lack of customer cooperation, lack of long-term strategy and planning); lack of top management attention (lack of top management commitment, lack of network designing, lack of infrastructure for distribution networks, lack of continuing research and study, lack of top management responsibility); lack of resources (lack of training courses and support, lack of financial resources, lack of enough incentives); lack of power quality involvement (lack of PQ measurement, lack of PQ consultants, lack of PQ standards, lack of PQ databases).

3 BARRIERS TO IMPLEMENTING A PQP

Since 1980, PQ issues have been causing real and significant disturbances to the distribution systems and end users worldwide, becoming a global concern, [28], [29], [30]. Hence, the lack of awareness of PQ could result in utilities still suffering from PQ problems caused by end users' sensitive equipment for industrial, agriculture, residential and commercial [14]. Therefore, providing sufficient introduction, definitions and explanations for the most widespread PQ terms, will help in identifying the more common PQ disturbances that occur. Moreover, those producing or using the power, in particular in less developed countries, should understand what PQ means.

The reason is that as long as the concept of PQ is misunderstood by both the staff of the electrical distribution company and the end users, then the severity of PQ issues will increase every day, because the demand for power will increase and even double [31].

Several authors and researchers have determined different aspects of barriers according to their experience and their studies on the implementation of PQP.

A study in the UK revealed eight major categories of PQP barriers: lack of staff awareness regarding PQ issues; lack of enough resources; lack of PQ training courses; lack of top management committed to implementing good PQP; lack of long-term strategy for successful implementation; lack of end users' awareness; lack of PQ standards and lack of regular maintenance [32].

A study conducted by Ghatol and Kushare found two aspects of PQP barriers in less developed countries; lack of network designing; and lack of end users' awareness regarding power quality [33]. A survey in the USA, conducted for the North American Delivery Systems found two barriers to PQP implementation; lack of customer cooperation i.e. illegal connection made by end users; and lack of top management responsibility to face customer complaints [16]. A study in a Massachusetts distribution system found three barriers to PQP implementation; lack of PQ standards; lack of cooperation by end users; and lack of management commitment regarding end users' complaints [34].

A study by EPRI in the USA pointed out nine components of PQP implementation barriers; lack of top management commitment, support and encouragement; lack of skills, knowledge and experience among engineers' and technicians; lack of proper teams to analyse PQ disturbances; lack of training courses; and lack of a PQ database [35]. Another study in the USA, Asia, Africa, Australia, South America and Europe revealed a lack of power quality awareness among end users; and lack of PQ training courses [17]. A further study in the USA revealed two barriers believed to hinder the successful implementation of PQP; lack of a utilities distribution structure; and lack of suitable management structure and operation [21].

A study in Malaysia found that five barriers to implementing a PQP were a ; lack of education programs; lack of PQ awareness and guidelines; lack of training courses and support; lack of continuing research and development; and lack of financial incentives to encourage the staff to resolve PQ issues [36].

A survey conducted in 8 developed European countries, namely; Austria, France, Italy, Poland, Portugal, Slovenia, Spain and the UK; found that a lack of end users' awareness; lack of employee awareness and skills; lack of management commitment; and lack of PQ measurements and maintenance are the main barriers to PQP implementation. These factors have led to huge economic losses in Europe, exceeding €150bn annually [37]. Another survey in Europe found that the main difficulties encountered during the implementation of PQP are lack of PQ awareness among top management, engineers and end users; lack of network designing, due to increased power demand; lack of PQ standards; lack of PQ measurement [38].

A study in Canada revealed that three main factors impede the wider spread of PQPs; lack of PQ consultants; lack of PQ standards; and lack of PQ awareness on the part of end users [2].

A study in the Netherlands found five significant difficulties in implementing PQP, namely; lack of a distribution networks infrastructure; failure to handle end users' complaints so as to identify the underlying problems;

lack of PQ contracts between suppliers and end users; increasing sensitive electronic equipments; lack of PQ training courses to raise the education and awareness levels of engineers to understand consumers' complaints better [39].

Another study in Germany found twelve barriers to PQP implementation; lack of distribution network designing, structure and size; lack of data on end users' load characteristics and structure; inadequate background and experience among employees regarding PQ; lack of PQ standards; lack of PQ measurement; lack of management planning and strategy [16].

A study in India found two major barriers to PQP implementation; lack of PQ measurement; lack of PQ awareness and skills among employees [40]. A second study in India found four significant categories of PQP barriers; lack of planning and designing the distribution network; lack of proper PQ teams; lack of PQ monitoring and databases to analyze customer complaints; and lack of PQ standards [41]. In Pakistan, a study found that lack of understanding PQ disturbances is a major obstacle to the implementation of a PQP to be achieved [42].

A study conducted by Moncrief, Dougherty, Richardson, and Craven found five main barriers to PQP implementation; lack of end users' awareness; lack of PQ equipment standards; lack of PQ awareness among employees; lack of PQ monitoring and databases regarding end users' complaints as a form of assistance to the utilities; lack of PQ measurements [43]. A study in Latin America found three barriers encountered during the implementation of PQP; lack of PQ monitoring and datasets; lack of PQ standards; lack of PQ employee' awareness and experience [44].

A study in Brazil found seven factors as the main barriers to PQP implementation; lack of distribution networks infrastructure; lack of studies and research; lack of distribution network design; lack of management planning; lack of technician and engineer skills and experience; lack of end users' awareness; lack of a clear strategy [45]. Table 1 shows the different and similar barriers by the above researchers.

TABLE 1

THE DIFFERENT AND SIMILAR OF POWER QUALITY PROGRAM DISCERNED BY THE ABOVE RESEARCHERS

Barriers	Country
lack of staff awareness, skills and experience	USA, European, India, Latin America, Brazil, Germany, Pakistan, Austria, France, Italy, Poland, Portugal, Slovenia, Spain and UK,
lack of enough resources	USA, UK
lack of top management commitment	USA, Massachusetts, Austria, France, Italy, Poland, Portugal, Slovenia, Spain and UK,
lack of long-term strategy and planning	USA, Brazil, Germany, UK
lack of end users awareness	USA, Asia, Africa, Australia, South America and Europe, Canada, Brazil, Austria, France, Italy, Poland, Portugal, Slovenia, Spain and UK,
lack of network designing	USA, European, India, Brazil, Germany
lack of training courses, and support	Malaysia, USA, Asia, Africa, Australia, South America and Europe, Netherlands, UK
lack of conducting research and studies	Malaysia, Brazil
lack of financial incentives	Malaysia
lack of customer cooperation	USA, Massachusetts
lack of top management responsibility	USA, Netherlands

lack of PQ standards	Massachusetts , European, Canada, India, Latin America, Germany, UK
lack of PQ measurement	India, USA, Germany, Austria, France, Italy, Poland, Portugal, Slovenia, Spain, UK,
lack of PQ consultants	Canada, India, USA
lack of DNS infrastructure	Netherlands, Brazil
lack of PQ monitoring and database	India, USA, Latin America
lack of regular maintenance	UK, Austria, France, Italy, Poland, Portugal, Slovenia, Spain

4 POWER QUALITY SURVEY IN A DISTRIBUTION NETWORK

Some of the power quality disturbances include: harmonics, short interruptions, long interruptions, voltage sags & swells, under voltage, over voltage, flicker & unbalance, transient & surge, low power factor and voltage collapse. These disturbances are considered in the statistical analysis presented in this paper. Results obtained from the survey indicate the current status of power quality supply in Libyan distribution utility staff's point of view. Almost 400 responses have been provided by one of the 3 major distribution networks, indicating opinions of the level of power quality among both residential, agriculture, commercial and industrial users.

The survey was mainly aiming:

1. To investigate the main reasons underlying power quality phenomena leading to power quality disturbances in Libyan distribution networks.
2. To identify the most significant factors, that would make a major impact on implementing power quality program in LDNs.
3. To develop a framework as guidelines through which LDNs could maintain and improve the power supply in terms of quality for its customers.

5 RESEARCH METHOD AND SURVEY INSTRUMENT

In Libyan distribution networks, empirical research is required to categorize and underline the barriers and benefits of PQP in the context of a distribution utility, which has not implemented power quality programs in the last two decades. The knowledge and results obtained from this study will guide Libyan distribution networks implementing PQP framework, including all departments and staff, who are directly responsible for remedying power quality disturbances, in tackling any power quality problems by setting up clear and long-term strategies, with crucial objectives and serious barriers. Therefore, the implementation of power quality program requires great attention from the top management to assist the distribution networks to achieve their goal of offering and providing a power quality program in practice [46]. The above literature review helps the researcher to understand the different barriers to PQP implementation and the expected benefits of PQP.

The questionnaire was sent to head managers, middle managers, engineers, technicians and employees, with total

number of 540 copies and it conducted in April-June 2009. Of 540 copies, 441 copies were returned, of which 397 were appropriate for data analysis, giving a response rate of 81%. The data were analyzed by using Statistical Package for Social Science (SPSS) software, version 15.0.1.1.

In addition, 44 face-to-face interviews conducted in this study to investigate why there were barriers to PQP implementation. The interviewees consisted of head managers, middle managers, engineers, technicians and employees from four departments, mainly those dealing directly with power quality issues. These were Planning, Training, Distribution, and Customer departments in LDNs. After the interviews conducted the data were transcribed and coded by using NVivo 9 [47]. Table 2 shows the 16 PQP Barriers.

TABLE 2

LIST OF POWER QUALITY PROGRAM BARRIERS

No	PQP Barriers
BA1	lack of staff awareness, skills and experience
BA2	lack of end users awareness
BA3	lack of customer cooperation
BA4	lack of long-term strategy and planning
BA5	lack of top management commitment
BA6	lack of network designing
BA7	lack of distribution networks infrastructure
BA8	lack of conducting research and studies
BA9	lack of top management responsibility
BA10	lack of training courses, and support
BA11	lack of financial resources
BA12	lack of enough incentives
BA13	lack of PQ measurement
BA14	lack of PQ consultants
BA15	lack of PQ standards
BA16	lack of PQ monitoring and database

Table 3 shows the type of distribution networks along with the categories of end users involved in the study. Large distribution networks were considered to have more categories of end users; the western distribution network (WDN1), southern-west distribution network (SWDN2) and eastern distribution network (EDN4); whereas small distribution networks had 1 to 2 categories of end users; the central distribution network (CDN3) and southern-east distribution network (SEDN5).

TABLE 3

TYPES OF DISTRIBUTION NETWORKS

Distribution Networks	Residential	Commercial	Industrial	Agricultural

WDN1	√	√	√	
SDN 2	√		√	√
CDN3	√	√		
EDN 4	√		√	√
SDN 5	√			√

6 POWER QUALITY SURVEY RESULTS

One of the points, which was clear, is that there was no power quality awareness program which can at least match the sudden growth in the economy. In response to this, there was significant need to conduct a power quality survey. It is revealed that lack of power quality awareness is the main issue. As a result of that, lack of awareness was found among the four main categories: Top management, Engineers, Technicians and End users. These are the people who suppose to solve Power Quality Disturbances (PQDs) or at least to be aware of the PQDs. These issues are seen as very crucial and fundamental requirements before start mitigation power quality disturbances.

6.1 Main Sources of PQ Disturbances

A summary of the main sources of the PQDs carried out by the survey is illustrated in Fig. 1.

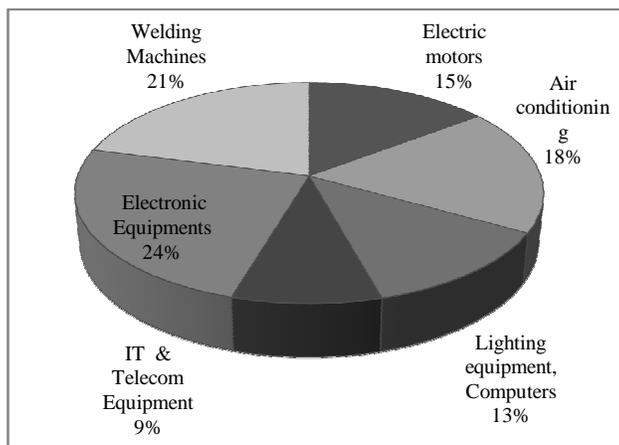


Fig. 1. Sources of PQ Disturbances - by responses %

The figure shows that electronics equipment is the largest source of PQDs. Although electronic equipments can be seen as a small load compare to, say, air conditioning loads; the disturbances generated by electronic equipments is extremely large with total harmonic distortion reaching 200% [31]. All loads shown in Fig.1 are non-linear loads and it is these types of loads which are in the increase.

6.2 PQ Disturbances

Table 4 illustrates the most common PQDs as seen by the people working within the surveyed power distribution network. It shows the industrial/commercial and residential figures as well as the total. In general the table shows that all

PQDs are taken seriously (they are all above 25%). In addition, table 4 indicates that around 79% of the participating respondents refer to Long Interruptions as one of the elements which causes PQDs, due to heavy loads such as the Artificial River Project and random private agriculture using large induction motors are connected to this network. Southern distribution network is only fed by one side of the transmission lines, which are driven far away from Alkhoms generation plants from the north to the south. As a result, the end users in this network are connected via different substations by transmission lines over a long distance, very far away from the generation source. For this reason, a 400 KV line was constructed and connected to this network to overcome the problems due to the long transmission distance. However, problems still persisted after the new line was introduced. It is also due to huge numbers of air conditioning units used, especially in summer. Many end users operate their air conditioning using "illegal" connections due to the un-satisfaction of the quality of electricity supplied. In addition, citizens started private agriculture projects, as water can be found at less than 30m below ground. They started cultivating the desert without consideration of the network capacity and the impact their activities would have on power quality. As a result, the network lacks sufficient efficiency and ability to provide good power quality to all sectors, including residential, artificial river, and private agriculture projects. All these projects appeared after 1999 economic blockade, led to major PQDs in the network.

TABLE 4

COMPARISON OF POWER QUALITY DISTURBANCES BETWEEN TWO DIFFERENT CATEGORIES

PQ Disturbances	Industrial & Commercial	Residential	Total
Harmonics	25%	26%	51%
Short Interruptions	29%	25%	54%
Long Interruptions	45%	34%	79%
Voltage Sags & Swells	44%	30%	74%
Under voltage	43%	26%	69%
Over Voltage	33%	27%	60%
Flicker & Unbalance	27%	25%	52%
Transient & Surge	30%	30%	60%
Voltage Collapse	32%	-	32%
Low Power Factor	36%	28%	64%

From table 4, this was clear evidence that both residential and industrial & commercial end users were affected due to long destination of transmission lines. It also due to a number of aspects being combined with varying user's categories and equipment categories in the network. Consequently, as shown in table 4, the unsatisfaction of the end users about poor power quality rose sharply since 1999.

6.3 Consequences of poor power quality

Table 5 shows the most important consequences appeared due to poor power quality between the two categories industrial and commercial, those close to 50% of all

consequences, which are defined as the most significant interrupts by the respondents. The major common consequences occurred in industry group are Transformers / cables overheating, Motors / process equipment damaged, Computers / electronics equipment damaged, and Relays & Contacts tripping, which effects on the operating process . The major elements different from the industry reported are Relays & Contacts tripping, Computers lock up, Computers / electronics equipment damaged , Data loss and Lights flicker for commercial.

TABLE 5
CONSEQUENCES OF POOR POWER QUALITY BETWEEN TWO DIFFERENT CATEGORIES

Consequences of poor power quality	Industrial	Commercial
Relays & Contacts tripping	36%	54.1%
Computers lock up	29%	56.6%
Computers / electronics equipment damaged	44%	43%
Data loss	32%	39%
Motors / process equipment damaged	54%	26%
Transformers / cables overheating	63%	29.5%
Lights flicker	27%	42.8%
Circuit breakers tripping	30%	19%

6.4 PQ Disturbances Measurement

Figure 2 illustrates the measurement history period for the power quality program in the past. It shows that approximately 56.6 % of respondents, who were surveyed, were not aware if there was PQP implemented to measure PQDs. This indicates that most of the engineers, technicians and head managers are not fully aware of power quality problems as a result of not being aware of the importance of it. Whereas approximately 26 % of respondents were knew about PQDs as a result of being aware of definitions of power quality.

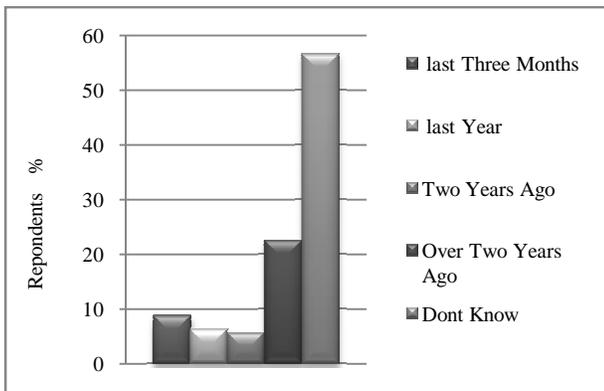


Fig. 2. PQP- by responses %

6.5 PQ Monitoring

Figure 3 illustrates the type of experts, who solved power quality problems in past measurements history.

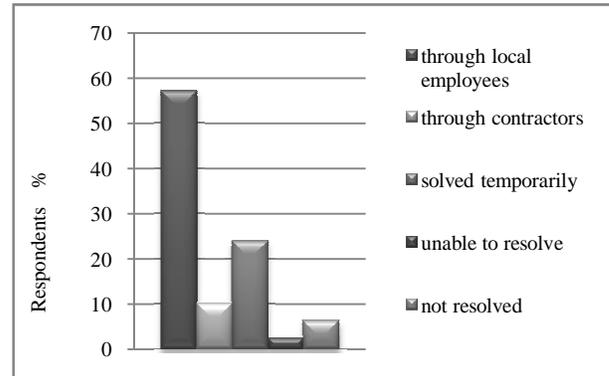


Fig. 3. PQ Monitoring - by responses %

Roughly 60 % of respondents identified that power quality problems were solved by local engineers and technicians, whereas 11 % of the respondents indicated that the problems were solved by contractors. From Fig 2, the engineers and technician surveyed predicted that they were aware of PQDs. It shows that the majority of the respondents (52.4%) comprising the educational qualification of high diploma degree, which considered as minimum education level. In order to deal with power quality events this level of education would felt to enable them to cope with the current level of severity of power quality. As a result, engineers and technicians are needed to be better educated and trained to be able to deal with power quality issues and found the urgent and appropriate solutions that decrease the disturbances. However, Figure 2 shows that almost 56.6 % of all the respondents were not aware of it in terms of definitions and problems.

6.6 Causes of PQ Disturbances

Figure 4 shows the most common group causing power quality problems. Lack of PQ awareness is considerably seen as the highest significant factor of causing the problems, where 31 % of the respondents cited that. As lack of awareness, approximately 26% of the end users connected illegally, as well as it increased the excessive use of electronic equipments, which introduced after 1999 of causing PQDs. This was due to non-linear equipments, which are very sensitive to power supply variation (long interruption 79%). In addition, lack of network designing at higher level as the third factor caused PQDs by 20 %. Therefore, the demand on the power generated has led the industries to demand and share it along with the increase demand in the domestic sector in the same line. These complex combinations required LDN to have power quality programs to make the network more efficient due to the complex interconnection [28]. Figure 4 gives more details about each aspect causes PQ issues with level of percentage by respondents.

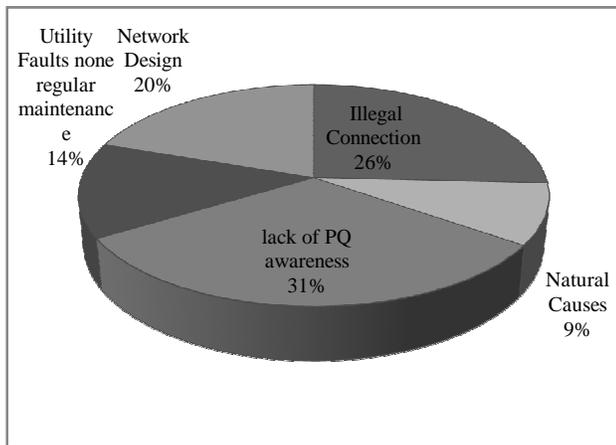


Fig. 4. Causes of PQ Disturbances - by responses %

7 PQ SURVEY FINDING AND DISCUSSION

Data gathered by the questionnaire from the distribution system respondents were checked in terms of accuracy, outliers and, normality; then analyzed using (SPSS) software. From the survey results, it appears that, there are significantly power quality issues occurring due to lack of awareness. From the results, there was a lack of awareness regarding the infrastructure for electricity supply between the Libyan government and General Electrical Company of Libya (GECOL) in term of designing LDN. Therefore, infrastructure was not at the level where the network could absorb the increases in demand due to population increases and requirements. Hence, it is one of the real reasons of causing PQDs in LDN, which has not met and adapted to the growing demand and the increase in economic growth. This is due to its cities, villages and remote areas with small populations located far away from each other.

Customer categorization can be of assistance in resolving power quality disturbances if the LDN adopt it not only in urban areas, but also in rural areas, where many villages and remote areas with small populations suffering bad service of power quality. In addition, distribution utilities should be accommodated with varied levels of consumers. As a result, each level of users can be determined separately and easily to diagnose and resolve any issues, which lead to PQDs.

On the other hand, the industrial, commercial, residential and agriculture sectors were found to be one of the biggest users causing PQDs. This has increased power quality problems, and the disturbances are generated constantly due to lack of awareness of the importance of it.

In addition, another cause of power quality problems is highlighted to be among the employees in head managers, engineers and technicians who were not aware of these issues to deal with its sources in increasing the problems as described in Fig. 2. As a result of that both end users complaints and attitudes are raised and caused significant reactions by faulty connect to the distribution networks, which is impacted on the quality of electricity. Thus, both power

quality guideline and datasets are required. This database is needed to record PQDs due to enormous use of new technology in electronic equipments, and in addition the framework is needed as guideline to increase the level of awareness for LDNs, including management, employees and end users in order to implement POP [48].

7.1 PQP Barriers

The Varimax - rotated factor matrix summarized in table 6, which accounted for about 64 % of the total variation. The correlation matrix disclosed the presence of many items <0.5 and items higher were considered to be important. Questions BA 1-4 belong to factor 1 and can be categorized under 'lack of awareness', whereas questions BA 5-9, belong to factor 2 and are categorized as 'lack of top management attention'. Questions BA 10-12 belong to factor 3 and pertain to 'lack of resources' and finally questions BA 13-16 belong to factor 4, dealing with 'lack of power quality involvement'.

The data were measured in order to evaluate the correlations between the barriers to PQP; therefore factor analysis was performed. The Kaiser-Meyer-Oklin (KMO) measure of sampling Adequacy value was 0.82, which exceeds the recommended value of 0.6 [49] and the Bartlett's Test of Sphericity was also highly significant (Chi-Square = 4847.51 with 561 degrees of freedom, at $p < 0.001$), reaching statistical significance in the correlation matrix. This implies that the factor analysis of 16 factors of PQP barriers was appropriate and confirms that all the items were statistically significant, which are judged to be an excellent validation of factor analysis.

TABLE 6
THE VARIMAX ROTATED FACTOR MATRIX

Items	Factor 1	Factor 2	Factor 3	Factor 4	Sig
BA1	0.797				0.035
BA2	0.731				0.033
BA3	0.699				0.337
BA4	0.666				0.036
BA5		0.801			0.044
BA6		0.754			0.049
BA7		0.676			0.021
BA8		0.641			0.447
BA9		0.623			0.043
BA10			0.837		0.022
BA11			0.787		0.044
BA12			0.755		0.242
BA13				0.766	0.031
BA14				0.711	0.041
BA15				0.701	0.029
BA16				0.671	0.128

The reliability test of Cronbach's α for all factors in questionnaire is 0.82. Cronbach's α above the cited minimums of 0.70 [50] is considered to be high and acceptable alpha, giving an evidence that the total Cronbach's alpha was judged

to be reliable for the questionnaire.

The ANOVA test results shown in table 6, out of 16 barriers, 12 were statistically significant different at the P value <0.05. The significant barriers are BA1, lack of staff awareness, skills and experience, BA2, lack of end users awareness, BA4, lack of long-term strategy and planning, BA5, lack of top management commitment, BA6, lack of network designing, BA7, lack of distribution networks infrastructure, BA9, lack of top management responsibility, BA10 lack of training courses and support, BA11, lack of financial resources, BA13, lack of PQ measurement, BA14, lack of PQ consultants, BA15, lack of PQ standards, and BA16, lack of PQ monitoring and database.

In addition, a post hoc Least Significance Difference (LSD) test was carried for these twelve barriers. The test found that large distribution networks WDN1, SDN2 and EDN4 faced some particular barriers compared to other small distribution networks in LDNs. SDN2 faces three factors; F1, lack of awareness, F2, lack of top management attention, and F4, lack of PQ involvement, whereas WDN1 and EDN4 face F1, lack of awareness, F4, lack of PQ involvement and F3, lack of resources. As a result, it can be referred that Libya's distribution systems have so far struggled to implement PQP effectively.

7.2 Interview Results

Table 7 shows the twelve difficulties discussed in the interviews, which are similar to what were obtained from the questionnaire. These results indicate that LDNs have not implemented PQP. It showed that the top management has not paid enough attention, support, commitment and responsibility to setting up long-term strategies to implement PQP. Therefore, LDNs have lost LD 464 million annually due to poor power quality and the failure to implement PQP [51].

TABLE 7

BARRIERS TO PQP FROM ANALYSIS OF THE INTERVIEWS

Barriers	Head Managers	Middle Managers	Engineers	Technicians
BA1	4.5%	69.85%	12.64%	13.01%
BA2	2.85%	56.26%	20.38%	20.51%
BA4	3.9%	60%	17.18%	18.92%
BA5	7.56%	56.68%	17.91%	17.84%
BA6	2.32%	71.44%	16.12%	10.12%
BA7	17.64%	50.1%	16.93%	15.33%
BA9	6.12%	76.75%	17.13%	0%
BA10	16.53%	44.35%	3.72%	35.4%
BA11	2.53%	58.26%	7.5%	31.71%
BA13	0%	95.27%	0%	4.73%
BA15	3.08%	83.28%	2.91%	10.73%
BA16	8.81%	64.18%	16.67%	10.34%

In addition, most of members of staff involved in improving PQDs are middle managers, 52.4% of who held of high diploma qualifications, which is considered the

minimum educational level. This means that they are not highly knowledgeable and aware enough to cope with the current severe level of power quality as well; moreover, this level of education would not enable them to understand and participate in implementing PQP. Almost 38% of engineers and technicians have between 6 and 15 years of experience, but lack awareness and skills. They should be better taught and trained before they can deal with PQDs improving.

8 PROPOSED PQP FRAMEWORK

Multivariable Linear Regression (MVLRL) was conducted to identify which factors have significant impact on PQP implementation [52]. An acceptable model was developed on the basis of these factors. It is clear that all these factors are significantly correlated, since all p values are less <0.05 and are substantially affected by the lack of awareness of the implementation of PQP in Libyan distribution networks as shown in Fig. 5.

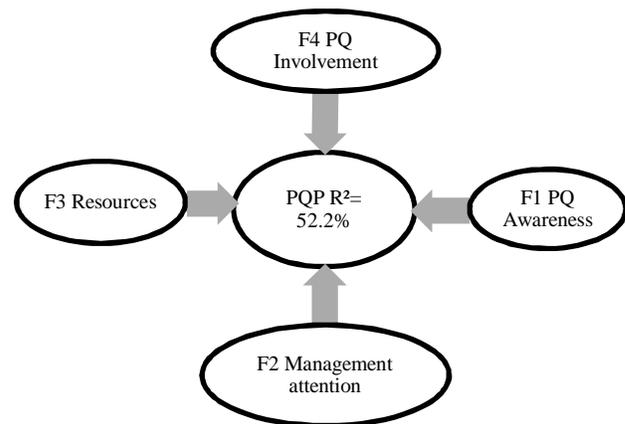


Fig. 5. Power Quality Program Model for LDNs

Table 8 shows the value of R² as 52.2% for this model, which indicates how much of the variability in the outcome is explained by the predictors. This also indicates that the validity of this model is very good. Consequently, this model can be accepted and applied for LDN to implement PQP, since all the predictors increase by one unit (see β value). It also indicates that the two factors most highly affected by lack of PQ awareness are F2 (β=34.5%) and F3 (β=31.6%). As a result, the regression analysis shows that the linear relationship between the outcomes, which is PQP, is explained by the model and predictor factors.

TABLE 8

REGRESSION RESULTS OF POWER QUALITY PROGRAM FACTORS

Scale	β	Std .Error	t	P	R ²	Cronbach's alpha
F1	0.202	0.031	4.538	<.001	0.522	0.811
F2	0.345	0.041	7.573	<.001		0.841
F3	0.316	0.029	8.097	<.001		0.806
F4	0.171	0.028	4.427	<.001		0.851

8.1 A Roadmap for Power Quality Program framework

The PQP framework is consisted three essential phases. Phase one designed to increase the awareness level. Phase two is involved in preparation of PQP, which contains seven crucial requirements. Phase three is designed to prevent the outstanding problems from phase 1 and 2 of not reoccurring again to determine both the weaknesses and obstacles facing the implementation, to reach a high level of power quality. Therefore, PQP framework is influenced by top management’s awareness, which must move from studies and recommendation to practice. This framework is designed as a guideline for implementation of PQP in the Libyan distribution networks environment. The progress of this framework and moving through from phase to other will be depended on the level of awareness, knowledge, skills gained respectively after each phase performed. Fig. 6 shows the three phases of the proposed model PQP of framework.

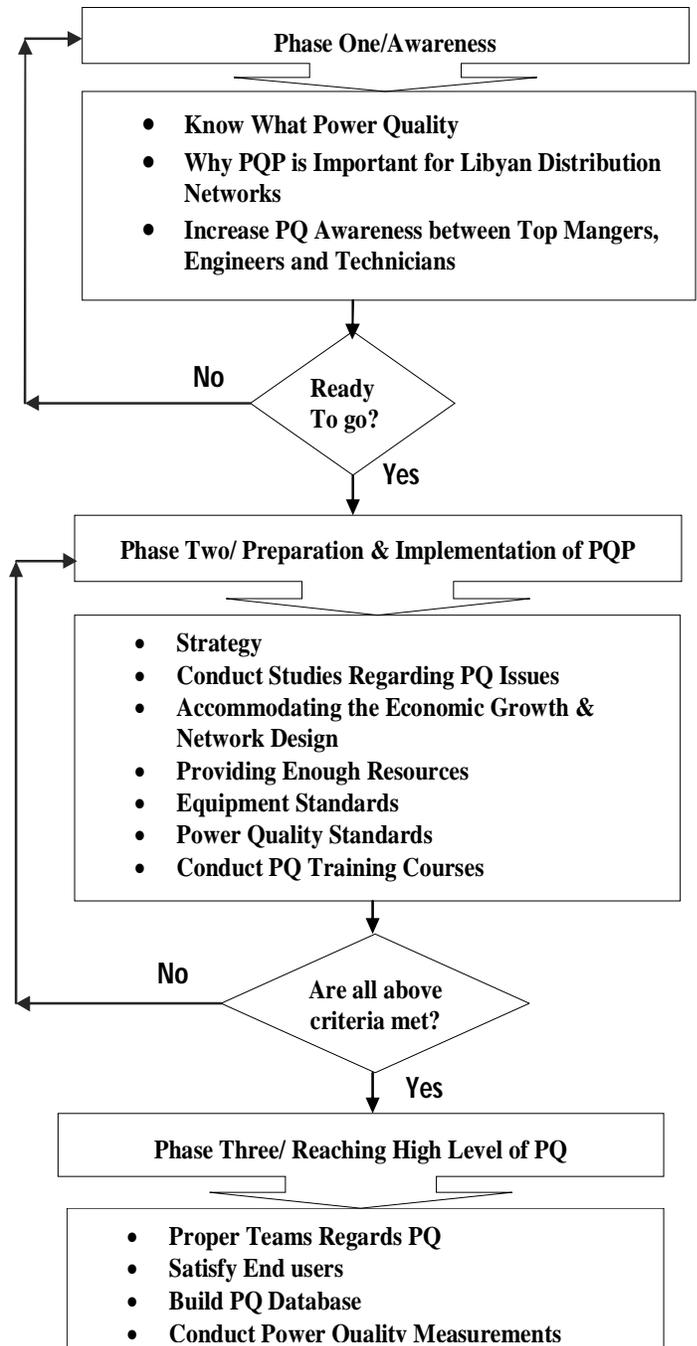


Fig. 6 Flowchart of PQP framework

8.2 Phase One Power Quality Awareness

Phase one is the fundamental of PQP. The purpose of this stage is to increase the awareness level regarding power quality program. The top management must be endorsed by paid more attention to reach the high level of understanding to prepare clear objectives , along with clear strategy for successful implementation of power quality program. This phase a waking up the top management to be eager and enthusiastic to start performance the program based on the staff level of both knowledge and awareness after understand the importance of PQP and its features. The expected output of phase one is shows in Fig. 7.

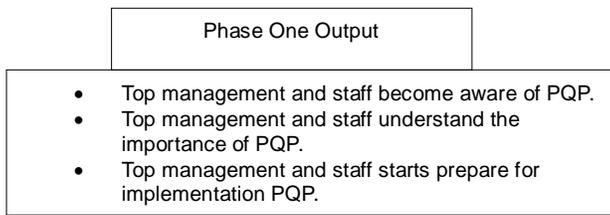


Fig. 7. The output of phase one

8.3 Phase Two Preparation & Implementation of PQP

Phase two is in conjunction with phase one. After the top management becomes aware and understands the importance of PQP, then this phase will enhance the distribution networks to start the implementation for PQP. The most important requirements of this stage is that the top management must set clear and long term strategic to continue built PQP to become one of Libyan distribution networks culture. One of the top management responsibilities is to develop and provide the needs of this phase to reach the high level of power quality program cross its networks. This step symbolizes the most critical factors of this framework and requires both top management commitment and employee's participation for PQP implementation in Libyan distribution networks as success factors. The expected output of phase two is shows in

No	DN1	DN2	DN3	DN4	DN5	Overall
BN1	3.84	3.96	3.45	3.27	3.8	3.66
BN2	3.91	3.56	3.54	3.73	3.53	3.65
BN3	3.65	3.68	3.54	3.64	3.4	3.58
BN4	3.51	3.52	3.68	3.51	3.47	3.53
BN5	3.48	3.48	3.82	3.53	3.33	3.52
BN6	3.73	3.56	3.67	3.49	3.46	3.58
BN7	3.52	3.48	3.49	3.55	3.66	3.54
BN8	3.76	3.88	3.82	3.77	3.93	3.83
BN9	4.25	3.31	3.75	3.53	3.48	3.66
BN10	3.43	3.68	3.73	3.25	3.52	3.52
BN11	3.48	3.66	3.61	3.52	3.56	3.56

Fig. 8.

Fig. 8. The output of Phase Two

8.4 Phase Three Reaching High Level of PQ

Phase three would facilitate Libyan distribution networks determine both the weaknesses and obstacles facing the implementation of power quality program. The previous two phases are designed to prevent the outstanding problems of not reoccurring again. The goal of implementing PQP framework is to increase the awareness level to practice and perform PQP framework practically with great attention from the top management. It also Focus on end users needs and satisfy them by considering their complaints in everyday process. Power quality improvements should be conducted by proper teams to measure and analysis power quality disturbances by building PQ database to monitor, measure,

analysis and compare it to PQ standards. Fig. 9 shows the output of phase three.

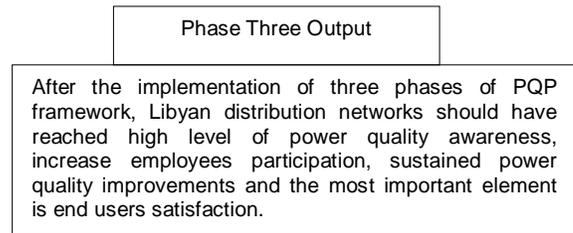


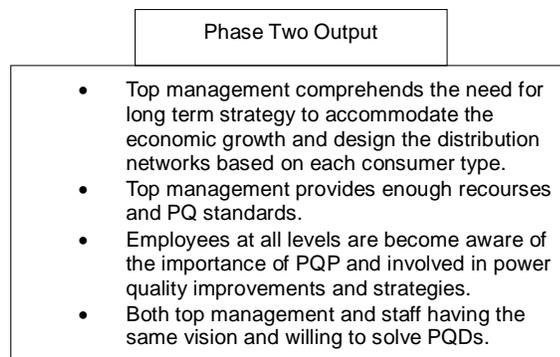
Fig. 9. The output of Phase Three

8.5 PQP Benefits

The overall results of PQP benefits are presented in table 9, which would have a positive impact on increasing end users awareness, increasing their satisfaction, improving PQ performance, reducing end users' complaints, monitoring and measuring PQ disturbances, providing PQ diagnosis systems and databases, reducing the huge losses through PQ costs, increasing top management awareness, increasing the employees' skills and awareness, increasing PQ training courses and providing strategic planning in LDNs.

TABLE 9

LIST OF POWER QUALITY PROGRAM BENEFITS OF SURVEY RESULTS



9 CONCLUSION

Significant power quality surveys and studies were critically reviewed to determine the critical factors regarding the implementation of PQPs. A power quality survey was conducted in LDNs west, east and south, networks as example one of distribution utilities in less developed countries. The survey provided various conclusions about occurrence of PQ issues, their sources and equipment affected LDNs. The results showed that most power quality issues were due to lack of PQ awareness and knowledge on part of Libyan distribution staff and customers. The rapid economic growth was a very significant factor, which has caused huge PQDs in LDNs after 1999. Statistical data also show that in the last two decades, LDNs have not implemented power quality program. This is mainly because there is no power quality department established yet, to influence the measurement of

power quality disturbances (PQDs). This absence of a power quality department is due to lack of awareness on the part of top management regarding the importance of power quality. This study also reveals poor implementation of PQP in LDNs. As a result, lack of power quality awareness has led LDNs to face twelve significant difficulties through not implementing PQP. According to qualitative analysis, this gap will continue if PQP barriers are not tackled.

One of the main challenges in implementing PQP is to link all the difficulties with both its objectives and strategies set by all departments. Hence, the implementation difficulties should be regularly assessed to identify the hidden reasons associated and causing poor implementation. Thus, without adequate knowledge, awareness, planning, designing, preparation, training, power quality standards, clear strategy, and most important the support of top management for this program, power quality disturbances will never end and their severity will affect all consumers.

The large distribution networks WDN1, SDN2 and EDN4 faced some particular barriers, unlike the smaller distribution networks in LDNs. SDN2 faces three factors F1, lack of awareness, F2, lack of top management attention, and F4, lack of PQ involvement; whereas WDN1 and EDN4 face F1, lack of awareness, F4 lack of PQ involvement and F3 lack of resources. The result of this is that Libya's distribution systems have struggled so far to implement PQP effectively. In general, the finding shows that LDNs suffer the four factors of PQP barriers. These four factors appeared in USA, European, India, Malaysia, Latin America, Brazil, Germany, Pakistan, Austria, France, Italy, Poland, Portugal, Slovenia, Spain and UK.

Approximately 75 % of the interviewees stated that there was no power quality program implemented in the past. From the questionnaire only 26 % of engineers, technicians and head managers were aware of power quality problems, while 56.6% were not aware. In response to this, new model for PQP framework is developed and proposed for LDN. The regression was sufficiently representative to conclude that the relationship between the model and the depended variables of power quality awareness is very strong and not accident. PQP implementation was found to have 11 benefits, which positively will impact on LDNs to improve power quality disturbances. They are needed because LDNs have not yet implemented PQP due to the failure to establish power quality departments.

REFERENCES

- [1] B.W. Kenedy, *Power Quality Primer*. USA: McGraw-Hill, 2000.
- [2] N. Labricciosa. A case study of power quality enhancement in an electrical distribution system. 1996.
- [3] R.C. Dugan, M.F. Mcgrannahan, S. Santoso, H.W. Beaty. *Electrical Power System Quality (Second ed.)* 1996.
- [4] E. I. Ortiz-Rivera, "Study of power quality: Problems in commercial buildings in Puerto Rico," in *Electric Utility Deregulation, Restructuring and Power Technologies*, 2004. (DRPT 2004). Proceedings of the 2004 IEEE International Conference on, 2004, pp. 301-305 Vol.1.

- [5] M. H. J. Bollen, "What is power quality?" *Electrical Power System*, vol. 66, pp. 5-14, 2003.
- [6] T. K. K. Forsten. *Security, quality, reliability and availability: Metrics definition: Progress report*. EPRI. USA. 2005.
- [7] M. McGranaghan, D. Brooks, D. Mueller. *EPRI reliability benchmarking application guide for Utility/Customer PQ indices: Guide for utilizing the EPRI reliability benchmarking, methodology power quality indices and customer specific indices as service quality benchmarks*. EPRI. USA. 1999.
- [8] S. S. M. R.S. Vedam. *Power Quality VAR Compensation in Power Systems (1st ed.)* 2009.
- [9] M.H.J. Bollen, *Understanding Power Quality Problems "Voltage Sags & Interruption"*. USA: John Wiley & Sons Inc, 2000.
- [10] D. Kottick. *Electrical power quality & utilization magazine, power quality monitoring System—Voltage dips, short interruptions and flicker*. 3(2), pp. 2. 2008.
- [11] W.M. Grady, A.H. Noyola, D. Dorr, P. Nelsen. *Results of power quality surveys in the united states: End user and utility perspectives*. Presented at *Power Quality Business Analyzer Survey Progress Report on PQ Costs*. 2000, [EPRI].
- [12] I. Chung, D. Won, J. Kim, S. Ahn and S. Moon, "Development of a network-based power quality diagnosis system," *Electr. Power Syst. Res.*, vol. 77, pp. 1086-1094, 6, 2007.
- [13] T. K. Abdel-Galil, E. F. Ei-Saadany and M. M. A. Salama, "Effect of new deregulation policy on power quality monitoring and mitigation techniques," in *Transmission and Distribution Conference and Exposition, 2001 IEEE/PES*, 2001, pp. 554-560 vol.1.
- [14] J. V. Milanovic and M. Negnevitsky, "Power quality problems and solutions: Current understanding," in *Harmonics and Quality of Power*, 1998. Proceedings. 8th International Conference on, 1998, pp. 30-35 vol.1.
- [15] L. Freeman, M. Marshall, R. Walling and J. Weiss, "Loss characteristics of local-delivery distribution systems," in *Electricity Distribution - Part 1*, 2009. CIRED 2009. 20th International Conference and Exhibition on, 2009, pp. 1-4.
- [16] J. Meyer, P. Schegner, G. Winkler, M. Muhlwitz and L. Schulze, "Efficient method for power quality surveying in distribution networks," in *Electricity Distribution, 2005. CIRED 2005. 18th International Conference and Exhibition on, 2005*, pp. 1-4.
- [17] A. Barnard, A. Van Voorhis. *Investigation utility-based power quality programs: Domestic and international perspectives*. EPRI. USA. 2000.
- [18] B.B. Angelo, *Handbook of Power Quality*. Italy: John Wiley & Sons Ltd, 2008.
- [19] M. R. Wadi, M. F. Bara, O. Carlson and K. M. Elarroudi, "Voltage stability analysis for the south-west Libyan electrical power system: Problem simulation and analysis," in *Universities Power Engineering Conference (UPEC), 2009 Proceedings of the 44th International, 2009*, pp. 1-5.
- [20] O. S. Al-Abaidi, A. B. Bendardaf, "Upgrading the local area network of the 220kV libyan eastern network control center," in *Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES*, 2005, pp. 1-4.
- [21] J. G. D. L.A. Schienbein. *Distributed energy resources, power quality and reliability –Background*. PACIFIC NORTHWEST NATIONAL LABORATORY. USA. 2002.
- [22] A. H. Hadod. *Libyan electrical power sector*. GECOL. Tripoli, Libya. 2009.
- [23] B. T. M. Elobadi, "Power quality report: Voltage profile," GECOL, Tripoli, Libya, 2009.
- [24] General Electrical Company of Libya, "Annual report," GECOL, Tripoli, Libya, 2009.
- [25] G. Nicholson, "Investigation of Data Reporting Techniques and analysis of Continuous Power Quality Data in the Vector Distribution Networks," 2006.

- [26] D.F. Warne. Newness Electrical Power Engineer's Handbook (2nd ed.) 2005.
- [27] M. A. Mousa, I. M. Saleh Ibrahim and I. M. Molokhia, "Comparative study in supplying electrical energy to small remote loads in Libya," *Renewable Energy*, vol. 14, pp. 135-140, 8, 1998.
- [28] J. Arrillaga, M. H. J. Bollen and N. R. Watson, "Power quality following deregulation," *Proceedings of the IEEE*, vol. 88, pp. 246-261, 2000.
- [29] O. GUL, "An Assessment of Power Quality and Electricity Consumer's Rights in Restructured Electricity Market in Turkey," *Istanbul Technical University, Istanbul*, vol. XIV, pp. 29-34, 2008.
- [30] Z. J. Paracha, S. A. Qureshi, "Application OF Power Quality Management Techniques and Methods In Power System of WAPDA PAKISTAN," vol. 4, pp. 1813-419, 2007.
- [31] B. E. Kushare, A. A. Ghatol, "Power quality characteristics of industrial electric distribution systems," *Institute of Engineering Education & Research, Nashik, India*, 2008.
- [32] Ph. Ferracci, "Power quality," *Cahier Technique Schneider Electric, UK, Tech. Rep. 199*, 2001.
- [33] G. D. N. Aniruddh. Technical report on power quality management. *Electrical and Electronics ENGG. USA*. 2003.
- [34] B. H. S. McNulty. Power quality problems and renewable energy solutions, analysis of the prospects for renewable PQ solutions in massachusetts. *Massachusetts Renewable Energy Trust. USA*. 2002.
- [35] C. M. D. Dorr. Power quality for satisfied commercial and residential customers field test plan: Monitoring residential power quality. *EPRI. USA*. 2000.
- [36] M. A. Hannan, A. Hussain, A. Mohamed and R. A. Begum, "Methodological Framework for Industrial Flicker Severity Assessment and Awareness," *Australian Journal of Basic & Applied Sciences*, vol. 4, pp. 5578-5583, 11, 2010.
- [37] J. M. R. Targosz, "Pan-european power quality survey," in *9th International Conference on Electrical Power Quality and Utilisation, EPQU 2007*, 2007, pp. 1-6.
- [38] H. De. Keulenaer, "The Hidden Cost of poor Power Quality," pp. Leonardo Energy, *European Copper Institute*, 2003.
- [39] W. T. J. Hulshorst, E. L. M. Smeets, and J. A. Wolse, "Premium power quality contracts and labeling," *Work package 2 of the Quality of Supply and Regulation project, Netherlands, Tech. Rep. 30620162-Consulting 07-0401*, 2007.
- [40] U. K. Deshpande, P. A. Chitre, "Total power quality management in a large distribution system: A case study national seminar on "Total power quality management"," *Mumbai, India*, 2009.
- [41] N. Rajmani, "ELECTRIC POWER QUALITY: SOME TECHNICAL AND COST PERSPECTIVES " *Global Energy Consulting Engineers, INDIA, Tech. Rep. 265*, 2008.
- [42] S. A. Qureshi, F. Mahmood, Z. J. Paracha, S. A. Kashif, "POWER QUALITY DISTURBANCES IN POWER SYSTEMS AND COMPARISON OF DEVELOPED COUNTRIES STANDARDS," *Electrical Engineering Department, UET Lahore, Pakistan*, vol. 4, 2007.
- [43] W. Moncrief, J. Dougherty, L. Richardson, K. Craven, "Power quality applications guide for architects and engineers," *EPRICSG, Palo Alto, CA, USA, Tech. Rep. 113874*, 1999.
- [44] B. C. A. Abreu, "Power quality site survey process for regulated electricity market," in *Transmission & Distribution Conference and Exposition: Latin America, 2006. TDC '06. IEEE/PES, 2006*, pp. 1-6.
- [45] R. J. R. Gomes, D. O. C. Brasil and J. R. de Medeiros, "Power quality management issues over the brazilian transmission system," in *Harmonics and Quality of Power, 2002. 10th International Conference on, 2002*, pp. 27-32 vol.1.
- [46] W. E. Kazibwe, R. J. Ringlee, G. W. Woodzell and H. M. Sendaula, "Power quality: a review," *Computer Applications in Power, IEEE*, vol. 3, pp. 39-42, 1990.
- [47] D. Silverman, Ed., *Interpreting Qualitative Data*. Thousand Oaks, California: Sage, 2006.
- [48] A. Eberhard, Ed., *POWER QUALITY*. Croatia: InTech Janeza Trdine 9, 51000 Rijeka, 2011.
- [49] A. P. Field, Ed., *Discovering Statistics using SPSS: (and Sex, Drugs and Rock'n'Roll)*. London; Thousand Oaks: SAGE Publications, 2005.
- [50] J. C. Nunnally, I. H. Bemstein, *Psychometric Theory*. New York: McGraw-Hill, 1994.
- [51] B. A. Hamilton, "Strategic plan corporate objectives; business, functional unit plans, financial model and lessons learned," *GECOL, Tripoli, Libya*, 2006.
- [52] H. M. S. Chandana Herath, V. J. Gosbell, S. Perera and D. Stirling, "Power quality survey factor analysis using multivariable linear regression (MVLRL),"

in Harmonics and Quality of Power, 2008. ICHQP 2008. 13th International Conference on, 2008, pp. 1-5.

Authors' information

Saad S. Sultan is PhD Student with the Department of Electronic and Computing Engineering, Brunel University, Uxbridge, UB8 3PH, UK (e-mail: Saad.Sultan@brunel.ac.uk).

Mohamed K. Darwish is a Senior Lecturer in Power Electronics with the Department of Electronic and Computing Engineering, Brunel University, Uxbridge, UB8 3PH, UK. (e-mail: Mohamed.darwish@brunel.ac.uk).

Case Study on JAVA based IDS

Dr. Sameer Shrivastava

Abstract - Hacking and intrusion incidents are growing terrifyingly every year with the roll out of new technology. Nothing can hide in today's digitally connected world. One can be traced on DNS, NSlookup, Newsgroups, web site trawling, e-mail properties etc. In our project, we tried our hand on designing of an Intrusion Detection System (IDS) that shall implement pre-defined algorithms for identification of attacks over a network. Java programming language has been used for the development, JPCap shall be used to provide access to the winpcap. Online capture of packets shall be done on the network i.e., the one coming on the interface of the network. The systems that are directly or indirectly connected on the Internet can be secured by using the IDS designed.

Index Terms – CaseStudy, IDS, JAVA, Security.

1. INTRODUCTION

Almost all companies and institutions are worried about the security of their network. All the intruders are searching new ways to break the privacy of everyone. Even though there is development in the field of filtration of intrusion to the infrastructure of the network via the Internet but the network is not yet safe.

But IDS is a comparatively new technology for intrusion detection methods that came forward in recent years. To prepare and deal with the network attacks is the main role of Intrusion detection system.

Anyone if tries to break into or misuse the system then it is called an intrusion. Stealing confidential data or misusing your email system for spam both comes under the category of "misuse". The concept of Global village has taken its origin with the coming out of Internet and the World Wide Web. Any kind of information is virtually easy to achieve any on the internet. Networking computers and associated devices are an advantage for this, and are in rapid progress. The intruders and people who provide security to the systems in networks are in race with each other. Our project can be hosted on the client to assist the administrator in detection of several intrusion attacks and inform the owner of the system and also provide security by blocking the malicious users based on their IP addresses.

2. NETWORK INTRUSION

Network Intrusion is a planned attempt to enter a network breaking the security and confidentiality of the information present in the systems of the network, and the person carrying out this is called as an Intruder. The network administrator is believed to defend his network from such persons and this software can help his in his efforts.

Dr. Sameer Shrivastava is an Associate professor in the Department of Computer Science and Engineering, in Global Nature Care Sangathan group of institutions, Jabalpur, M.P. India. He received his Ph.D. (Computer Science and Engineering) from Bhagwant University, Ajmer, Rajasthan in 2011. He has 14+ years of experience in teaching and research. His areas of specialization include Network Security. He is a Cisco Certified Network Associate, SUN certified and Microsoft Certified Professional. He has published papers in many International and national level Journals on Network Security and Computer Networks

2.1 Intrusion detection systems (IDS)

A system responsible for detecting abnormal, inappropriate, or other data that may be considered illegal occurring on a network is an IDS. An IDS is subject to capture and inspect the traffic, despite of whether it's permitted or not. An alert is generated, based on the contents, at either the IP or application level.

2.2 Categories of Intrusion Detection System

IDS can be classified into three categories: signature based detection systems, anomaly based detection systems and specification based detection systems.

2.2.1 Signature Based Detection Systems

Signature based detection system are effective against known attacks, and get updated from new patterns or else will be ineffective in case of unknown previous threats or new releases.

2.2.2 Anomaly based Detection System

For the implementation of this system we must be aware of the normal behavior of the network, since it is based on rules or heuristics rather than patterns or signatures. Anomaly based detection system is able to detect previous unknown threats, but the false positive ration is quite high.

2.2.3 Specification based Detection System

It monitors the processes and matches the actual data with the program and in case of any abnormal behavior shall issue an alert. They must be maintained and updated whenever a change is there on the surveillance programs, in order to detect the previous attacks. The number of false positive ratio is comparatively less to anomaly detection system approach.

2.3 Classification of Intrusion Detection System

Intrusion detection system are classified into three types

2.3.1 Host based IDS

This is placed on either the server or workstation, where the data is analyzed locally and information is collected for this data from different sources. Anomaly and signature

based detection system can be used here.

2.3.2 Network based IDS

These are deployed on strategic point in network infrastructure. It is able to capture and analyze data to detect known attacks by either comparing patterns or signatures of the database, or by detecting illegal activities by scanning traffic for anomalous activity. We can also call them "packet-sniffers", since they capture the packets passing through the communication mediums.

2.3.3 Hybrid based IDS

This is a mix and match of both the techniques and provides logical complement to NID and HID.

3. NEED FOR AN IDS:

Intrusion detection devices form a vital part of any network. With a constant evolution of the Internet new vulnerabilities and exploits are found regularly, which provide an additional level of protection to notice the existence of an intruder, and help to provide liability for the attacker's action.

The need for IDS is critical due to identification of four different types of attacks.

3.1 Denial of service

Network-based such type of an attack is one of the easiest types. It requires very less effort to fully consume resources on the target computer, to starve the target computer of resources, or to cause critical services to fail or malfunction. Internal corporate networks usually do not have internal filtering barricades against common denial-of-service attacks, such as flooding.

3.2 Threat to Confidentiality

Some viruses are capable enough to get attached to existing files on the system they infect and then send out the infected files to others. This results in distribution of the confidential information without the author's permission.

3.3 Modification of contents

News sites modification, production of bogus press releases, and many other activities are done by Intruders, all of which could have economic impact.

3.4 Masquerade

When one entity pretends to be a different entity it's known as masquerade. After a valid authentication series has taken place it can be captured and replayed, thus enabling an authorized entity to obtain extra privileges which previously had less, by impersonating an entity that has those privileges.

A system with internet connection and provision of TCP-based network services (such as a Web server, FTP server, or mail server) is potentially subject to this attack. In addition to the attacks launched at explicit host, they could also be

initiated against your routers or other network server systems if these hosts enable (or turn on) other TCP services (e.g., echo).

The consequences of the attack may vary system to system; however, the attack itself is original to the TCP protocol used by all systems.

4. REQUIREMENTS FOR IDS

There are two levels of abstraction to list the requirement of IDS.

High Level Requirements:

- To develop a capable application that can sniff the traffic, to and from the host machine.
- Development of an application that is competent of analyzing the network traffic and detects numerous pre-defined intrusion attacks and mappings.
- Development of such an application that warns the owner of the host machine, about the likely occurrence of an intrusion attack and information is provided regarding that attack.
- Such an application to be developed that should block traffic to and from a machine that is identified to be potentially malicious and is defined by the owner of the host machine.

Low Level Requirements:

- To develop an application capable enough to display the incoming and outgoing traffic from the host machine in the form of packets to the owner of the host.
- An application that detects occurrence of Denial of Service attacks such as Smurf Attack and Syn-Flood Attack is required.
- Development of an application that detects attempts to map the network of the host, using techniques such as Efficient Mapping and Cerebral Mapping.
- Such an application is required that detects actions attempting to gain unauthorized access to the services provided by the host machine using techniques such as Port Scanning.
- To develop an application that maintains a "Log Record" of identified intrusion attacks done on the host in the present session and also displays it upon request.
- Activation or de-activation of each of the Attack Detection methods should be possible.
- To provide a selection procedure for the user of the host for framing Rules which explicitly specify the set of IP addresses to be blocked or allowed. These Rules shall determine the flow of traffic at the host.

5. USE AND SCOPE

Use of the system: The system must detect certain familiar

intrusion attacks on the host system and warnings to be displayed to the user and also store data regarding the IP addresses that allows the traffic based on the data.

Scope of the system: Based on the input provided by the user, the system frames certain rules. The traffic then flows to and fro based upon the rules. Some well-known attacks are also detected and warnings are given.

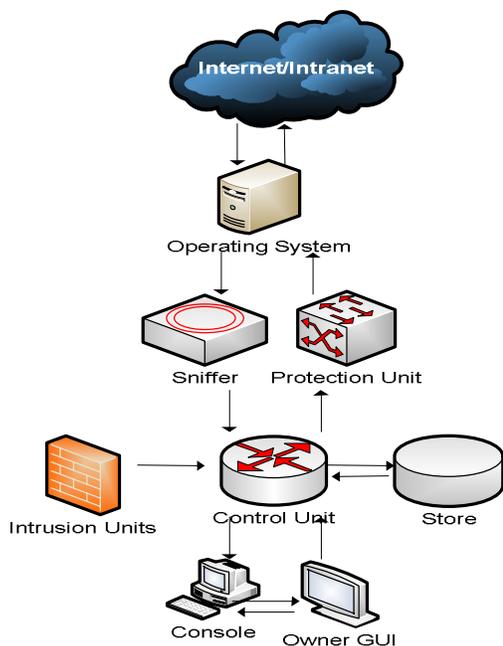
5.1 System overview

On the arrival of the packets, they are sniffed by the sniffer and then various processing techniques are applied to detect the attacks, and the users are warned against it. These are predefined and well known attacks. The following attacks detection are implemented in the system.

- Port Scanner,
- Smurf Attack,
- SYN Flood Attack,
- Efficient Mapping, and
- Cerebral Mapping.

The system has the provision of blocking traffic from a specific IP address which has been recognized to be malevolent or troublesome. Provision has been done to allow traffic from specific IP addresses for some trusted systems, whose traffic is not monitored. Unknown hosts traffic is monitored and any possible attacks are informed to the user.

6. SYSTEM STRUCTURING



Detailed Class Descriptions:

The system contains 31 classes of which 7 are inner classes. The following are the important classes identified, DataProcessor, PacketCapture, ControlUnit, IntrusionAttacks, Console, IntrusionUnit, Attack,

SimpleDES, Rule, ProtectionUnit, XmlData, Owner, and IDSMain.

Design with Reuse as goal

Almost all the components in the system are sequence independent from other components, which are designed in such a way that they can be reused by other systems. The components that can be used as Reusable components are as follows:

PacketCapture, XmlData, SimpleDES, IntrusionUnit, and Console.

The usage of JAVA in the development makes it Platform independent making the code portable on any Operating system.

7. CONCLUSION AND FUTURE WORK

All the systems present in the network and connected directly or indirectly to the Internet can be secured by basic detection techniques provided in IDS. Performing such a task goes hand in hand with success as well as failure in fulfillment of the objective. At least the job is done. But at last the Network Administrator has to make sure that his network is not in danger. This software does not completely defend network from Intruders, but the very purpose of IDS is to help the Network Administrator to track the bad guys on the Internet who are suppose to bring your network to a breach point and thus making it vulnerable to attacks. The following is an attempt to show what should be the source of action while using the software and after an attack has been detected by IDS. Like other conventional IDS this also provides facilities for Intrusion Protection. The blocking or allowing particular IP, range of IPs or a subnet IPs by applying relevant rule on the Operating system depends on this. This is a reusable system. Due to the high end flexibility and extensibility provided in the design of the system it becomes easy to add more number of attacks to the system in future.

Java has been the core for the development, thus making it platform independent, yet it has been tested only on WindowsXP. Though it will work fine on other operating systems also and thus satisfy the requirements and pre-requisites for the IDS system. A log is maintained that is valid only for the current session and contains no information about the past sessions, which can be enhanced in future. Enhancements in the system are possible on the works listed below:

The present system can display the log information but none of the techniques to analyze the data present in the log records and extract knowledge is present. Extension is possible by incorporating Data Mining techniques for analysis of the log records which may help in efficient decision making. The present system only detects the known attacks, which can be extended by incorporating Intelligence into it to gain knowledge by itself by analyzing the growing traffic and learning new Intrusion patterns. Presently system runs on an individual host machine and is not a distributed

application; while in future it can be extended to as a distributed application where different modules of the same system running on different machines may interact with each other thus providing distributed detection and protection the machines on which the system is running.

REFERENCES

- [1]. M. Cagalj, S. Capkun, and J.-P. Hubaux, "Wormhole-Based Anti-Jamming Techniques in Sensor Networks", in IEEE Transactions on Mobile Computing, January 2007
- [2]. Patwardhan, J. Parker, M. Iorga, A. Joshi, T. Karygiannis, Y. Yesha, Threshold-based intrusion detection in ad hoc networks and secure AODV, Ad Hoc Networks 6 (2008) 578–599.
- [3]. [42] S.A. Razak, S.M. Furnell, N.L. Clarke, P.J. Brooke, Friend-assisted intrusion detection and response mechanisms for mobile ad hoc networks, Ad Hoc Networks 6 (2008) 1151–1167
- [4]. N. Komninos, C. Douligeris, LIDF: layered intrusion detection framework for ad hoc networks, Ad Hoc Networks 7 (2009) 171–182.
- [5]. Idika, N. & Mathur P. A., "A Survey of Malware Detection Technique", In Proceeding of Software Engineering Research Center Conference, SERC-TR286, 2007.

Analysis and Future Approach of Ultra Wideband Technology

Manu Bali

Abstract— Ultra-wideband (UWB) technology is a revolutionary wireless technology used to transmit large amounts of digital data short distances (up to 230 feet) over a very wide bandwidth (from 1 gigahertz [GHz] up to 10 GHz [17]) and at very low power levels (less than 0.5 milliwatt). Unlike typical radio frequency broadcasts that use continuous sine waves to transmit data, UWB uses precisely positioned pulses at specific time intervals to transmit the signals across a wide spectrum. Ultra-Wideband (UWB) wireless is a rapidly growing technology that promises to revolutionize low power, short-range wireless applications. UWB has quickly emerged as the leading technology for applications like wireless Universal Serial Bus (USB) and short-range ground penetrating radars. UWB radios differ from conventional narrow-band radios, with a variety of specialized test demands. Enormous signal bandwidths, short duration pulses and transmit Power Spectral Densities (PSDs) [1] near the thermal noise floor, make UWB testing difficult. Fortunately, leading instruments like the Tektronix Arbitrary Waveform Generators (AWG), RFXpress waveform creation software and Digital Phosphor Oscilloscopes (DPO) with UWB measurement software offer solid solutions to UWB test challenges. In this Technical note we explain the concepts behind UWB technology, its unique hardware and software architectures, and future applications.

Index Terms— Arbitrary Waveform Generators, Digital Phosphor Oscilloscopes, Narrowband Communication, Power Spectral Density, Radio Free technology, Ultra Wideband, Universal Serial Bus.

1 INTRODUCTION

Ultra Wideband was traditionally accepted as [pulse radio](#), but the FCC and ITU-R now define UWB in terms of a transmission from an antenna for which the emitted signal bandwidth exceeds the lesser of 500 MHz or 20% of the center frequency. Ultra wideband (also known as UWB or as digital pulse wireless) is a wireless technology for transmitting large amounts of digital data over a wide spectrum of [frequency](#) bands with very low [power](#) for a short distance. UWB uses precisely positioned pulses at specific time intervals to transmit the signals across a wide spectrum [7].(Figure 1.)

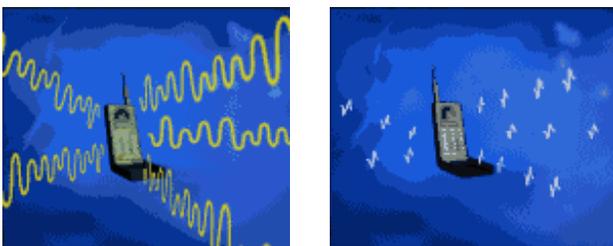


Figure1. Continuous Sine Waves vs. Time Modulated Pulses

Ultra wideband radio not only can carry a huge amount of data over a distance up to 230 feet at very low power (less than 0.5 milliwatts), but has the ability to carry signals through doors and other obstacles that tend to reflect signals at more limited bandwidths and a higher power. Ultra wideband can be compared with another short-distance wireless technology, [Bluetooth](#), which is a standard for connecting handheld

wireless devices with other similar devices and with desktop computers.

Ultra wideband broadcasts digital pulses that are timed very precisely on a carrier [signal](#) across a very wide spectrum (number of frequency channels) at the same time. Transmitter and receiver must be coordinated to send and receive pulses with an accuracy of trillionths of a second. On any given frequency band that may already be in use, the ultra wideband signal has less power than the normal and anticipated background [noise](#) so theoretically no interference is possible. Time Domain, a company applying to use the technology, uses a microchip manufactured by IBM to transmit 1.25 million bits per second, but says there is the potential for a data rate in the billions of bits per second.

2 CONCEPT

This concept doesn't stand for a definite standard of wireless communication (the standard is being developed now is still far from completion); this is a method of modulation and data transmission which can entirely change the wireless picture in the near future. Before going on future view let's take a look at the diagram that demonstrates the basic principle of the UWB:

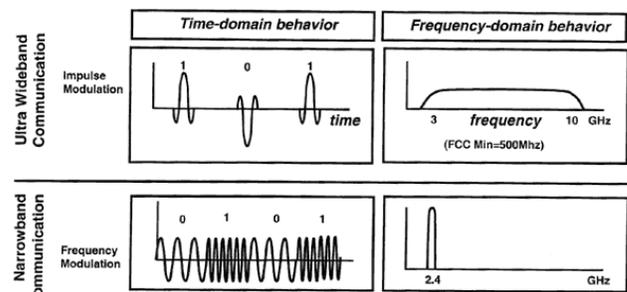


Figure 2: an approach on Narrowband and UWB Communication

- Manu Bali is currently Working as a lecturer in electronics and communication engineering in Maharishi Markandeshwar University, India, PH-918901163678. E-mail: manu4u2018@gmail.com
- Abhay Bindal is currently Working as a lecturer in electronics and communication engineering in Maharishi Markandeshwar University, India, PH-919996144047. E-mail: abhaybindal@gmail.com

The UWB is above and the traditional modulation is below which is called here Narrow Band (NB), as opposed to the Ultra Wideband. On the left we can see a signal on the time axis and on the right there is its frequency spectrum, i.e. energy distribution in the frequency band. The most modern standards of data transmission are NB standards - all of them work within a quite narrow frequency band allowing for just small deviations from the base (or carrier) frequency. Below on the right you can see a spectral energy distribution of a typical 802.11b transmitter. It has a very narrow (80 MHz for one channel) dedicated spectral band with the reference frequency of 2.4 GHz. Within this narrow band the transmitter emits a considerable amount of energy necessary for the following reliable reception within the designed range of distance (100 m for the 802.11b). The range is strictly defined by FCC and other regulatory bodies and requires licensing. Data are encoded and transferred using the method of frequency modulation (control of deviation from the base frequency) within the described channel.

Now take a look at the UWB - here the traditional approach is turned upside down. In the time space the transmitter emits short pulses of a special form which distributes all the energy of the pulse within the given, quite wide, spectral range (approximately from 3 GHz to 10 GHz). Data, in their turn, are encoded with polarity and mutual positions of pulses. With much total power delivered into the air and, therefore, a long distance of the reliable reception, the UWB signal doesn't exceed an extremely low value (much lower than that of the NB signals) in each given spectrum point (i.e. in each definite licensed frequency band). As a result, according to the respective FCC regulation, such signal becomes allowable although it also takes spectral parts used for other purposes:

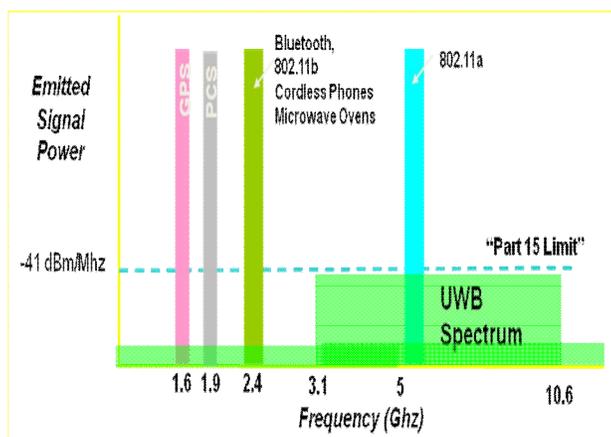


Figure 3: Different Standards in terms of Emitted Signal Power

So, the most part of energy of the UWB signal falls into the frequency range from 3.1 to 10.6 GHz, and the energy spectral density doesn't exceed the limit determined by the Part 15 of the FCC Regulations (-41dBm/MHz). Below 3.1 GHz the signal almost disappears, its level is lower than -60. The more ideal the form of a pulse formed with the transmitter, the less the energy goes out of the main range. But however that may be, the permissible deviation of the pulse from the ideal form

must be limited, hence the second purport. The spectral range lower than 3.1 GHz is avoided not to create problems for GPS systems whose accuracy of operation can suffer a lot from outside signals even if their density is lower than -41. That is why 20 dBm (up to -60) were reserved in addition at the spectral range up to 3.1 GHz; it is not obligatory but it seems to be welcomed by military bodies.

The total energy of the transmitter which can fit into this band is defined by the area of the spectral characteristic (see filled zones on the previous picture). In case of the UWB it's much greater compared to the traditional NB signals such as 802.11b or 802.11a [14]. So, with the UWB we can send data for longer distances, or send more data, especially if there are a lot of simultaneously working devices located close to each other.

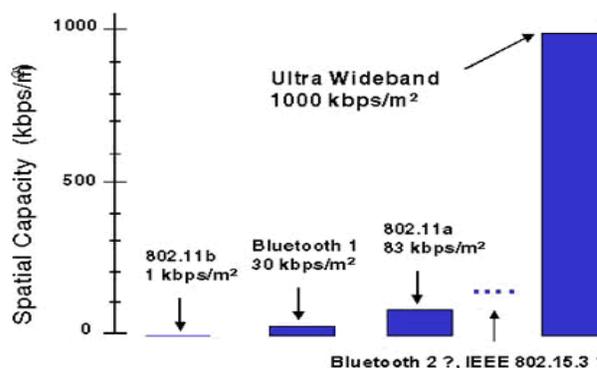


Figure 4: Spatial Capacity Comparison

A The UWB actually tries to solve the problem of inefficient spectrum utilization, like the Hyper Threading solves the problem of idle time of functional processor units. Frequency bands dedicated for different services remain unused for the most part of time - even in a very dense city environment - at each given point of time the most part of the spectrum is not used, that is why the radio spectrum is used irrationally:

1. Most frequencies are not used all the time. That is a low frequency effectiveness of the spectrum utilization.
2. Guard channels necessary for NB modulations (gaps between channels to eliminate pickups). That is a low frequency effectiveness of the spectrum utilization.
3. Excessive and, as a rule, uncontrolled power of transmission (and, therefore, transmission range) of signals even if a distance is quite short. That is a low spatial effectiveness of the spectrum utilization.

Whatever direction we are looking into nothing seems to be good - so, it's high time to start improving methods of radio communication and division of the air. In case of the NB a frequency and width of the dedicated spectral range for the most part (though the real situation is much more complicated) defines a bandwidth of the channel, and the transmitter's power defines a distance range. But in the UWB these two concepts interwine and we can distribute our capabilities between the distance range and bandwidth. Thus, at small distances, for example, in case of an interchip communication, we can get huge throughput levels without increasing the total trans-

ferred power and without cluttering up the air, i.e. other devices are not impeded. Look at how the throughput of data transferred in the UWB modulation depends on distance:

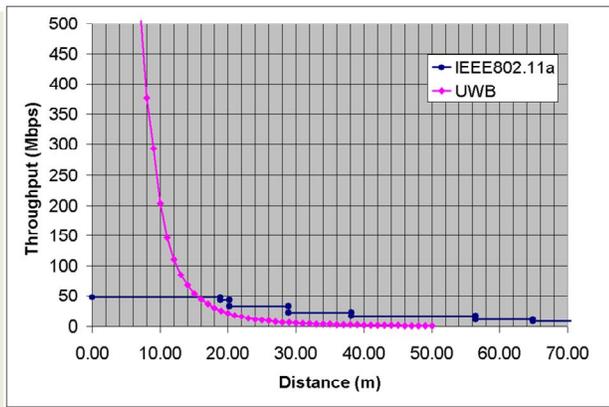


Figure 5: Throughput vs Distance

While the traditional NB standard 802.11a uses an artificially created dependence of throughput on distance (a fixed set of bandwidths discretely switched over as the distance increases), the UWB realizes this dependence in a much more natural way. At short distances its throughput is so great that it makes our dreams on the interchip communication real, but

at the longer distances the UWB loses to the NB standard. Why? On the one hand, a theoretical volume of the energy transferred, and therefore, the maximum amount of data, is higher. On the other hand, we must remember that in a real life information is always transferred in large excess. Beside the amount of energy, there is the design philosophy which also has an effect. For example, a character of modulation, i.e. how stably and losslessly it is received and detected by the receiver.

3 TRANSMITTER AND RECEIVER CONFIGURATIONS

The earliest radio transmissions by Marconi were UWB in the sense that Marconi's spark-gap transmitter in effect generated short pulses and occupied a relatively large bandwidth, but means for using spreading gain to enable multiple access were not available. Soon after the potential of radio as a medium for communication was understood, efficient methods for sharing the medium were sought and found that involved heterodyning and narrowband, tuneable transmitters and receivers. An example super heterodyne receiver diagram is given in Figure 6. that features double conversion to reject harmonic images of the signal that are unwanted by products of the heterodyning (multiplication) operations. With the proliferation of narrow-band wireless devices today and the continual development of new devices for the wireless market, the trend is for the transmitters and receivers to become smaller and simpler.

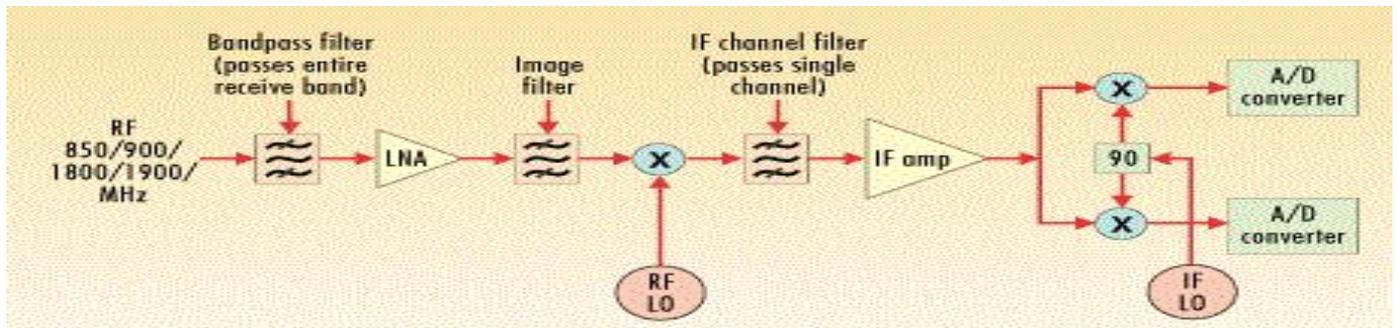


Figure 6: Double-conversion super heterodyne receiver

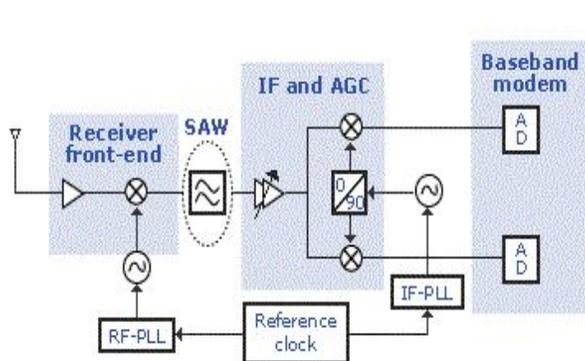


Figure 7: Typical Digital Heterodyne receiver

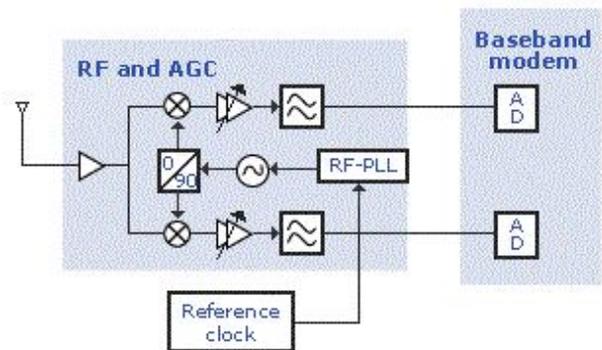


Figure 8: single-chip direct conversion receiver (right) that integrates RF and IF without a SAW filter



Figure 9: Concept of UWB baseband system implementation

For example, Figure 7 shows a typical digital heterodyne receiver using a surface acoustic wave (SAW) filter and a “one chip” receiver based on direct conversion to baseband that does not require the SAW filter. As such advances in digital processing became cheaper and more efficient; the use of UWB waveforms in radar and communication applications also has become feasible.

Ideally in carrierless (baseband) transmission, as illustrated in Figure 8, the radio system can operate without local oscillators and the sometimes complex filtering needed to control emissions and spurious radiations that accompany heterodyning. It is almost, but not quite, as simple as utilizing the same kind of transmissions as those that are unintentionally emitted by the printed circuit board of digital devices, with the antenna connecting directly to the integrated circuit containing the baseband processing logic.

4 UWB APPLICATIONS

UWB technology can enable a wide variety of WPAN applications. Examples include:

- Replacing cables between portable multimedia CE devices, such as camcorders, digital cameras, and portable MP3 players, with wireless connectivity
- Enabling high-speed wireless universal serial bus (WUSB) connectivity for PCs and PC peripherals, including printers, scanners, and external storage devices
- Replacing cables in next-generation Bluetooth Technology devices, such as 3G cell phones, as well as IP/UPnP-based connectivity for the next generation of IP-based PC/CE/ mobile devices
- creating ad-hoc high-bit-rate wireless connectivity for CE, PC, and mobile devices

5 A CHALLENGING TEST PROBLEM

UWB signals pose many challenging test and measurement issues that demand special test instrument capabilities. Generating and analyzing ultra broadband test signals for UWB requires high performance arbitrary waveform generators like the Tektronix AWG7000 series and very broadband digital phosphor oscilloscopes like the DPO70000 series that can support the enormous bandwidth requirements of the UWB signal. UWB signal requirements present broadband amplitude and phase flatness challenges. Transient UWB pulses can be distorted by the spectral amplitude and phase flatness from both the test signal generator and measurement instruments. Pulse distortion effects in turn alter the spectral properties of

UWB signals. For narrowband signals, test equipment is typically selected such that its bandwidth is significantly larger than the desired signal bandwidth to be measured, minimizing flatness issues. However, for UWB signals it is not possible to have a vastly wider test equipment bandwidth.

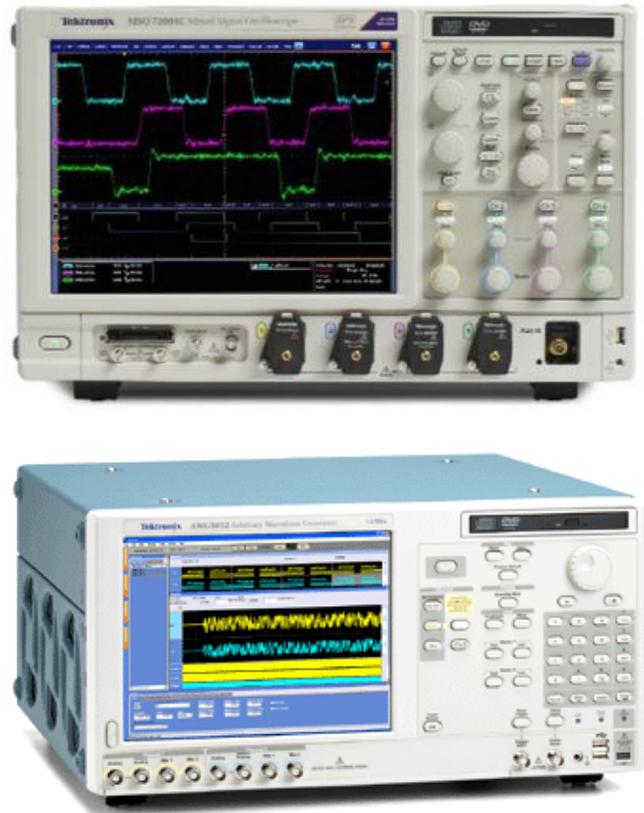


Figure 10: Typical Instruments like AWG7000 series and DPO70000 can generate and capture complex ultra-wideband signals.

Another problem encountered when testing UWB signals is the limited measurement bandwidth options available. Even simple power spectral density measurements can be difficult, as regulations require a 50 MHz resolution bandwidth (RBW) few spectrum analyzers support. Add to these challenges Time Frequency Codes (TFC) that spread the UWB signal, and device test can be a major challenge without the right test equipment. To understand which test solutions are appropriate for UWB, let us briefly review what UWB technology is all about and what makes up these fascinating signals.

6 TECHNOLOGY CONSIDERATIONS

For UWB technology to become a widely adopted radio solution, a few key areas need to be resolved:

- Interoperability
- Ease of product integration and certification
- Overall solution cost (to the OEM)
- Global spectrum allocation

Many Industries are addressing a number of these issues through investment strategies, research, participation in wired and wireless communications initiatives, and product development. Some are also developing protocols that will take full advantage of the strengths of UWB technology. The WUSB specification developed through the Wireless USB Working Group and the UPnP work done through the Digital Home Working Group (DHWG) is examples of industrial contributions.

7 THE FUTURE: RADIO FREE

UWB Technology envisions a future in which all devices are connected by smart radios. The vision is called Radio Free, and it embodies the concept of a smart radio that can reprogram and reconfigure itself based on available spectrum, the desired application, and the device at hand. Configurations would include an 802.11 radio for communicating with a WLAN hotspot, a Bluetooth Technology radio for communication with a cell phone, or a UWB radio for participation in a WPAN. To promote this vision, it is involved in all areas of the RF space. In wireless wide area networks (WWAN), it could be the case of WiMAX. Now, with support of UWB technology for the WPAN space, the concept of Radio Free technology is one step closer to reality.

8 CONCLUSION

UWB and the associated networking protocol efforts are in the early stages of development, and several key deployment scenarios are being defined and evaluated. UWB complements currently deployed wireless networks in the WLAN environment, plus it extends high bit-rate, multimedia connectivity to WPANs supporting PC, CE and cellular devices. This combination will enable true convergence of computers, consumer electronics and mobile communications. A common radio platform that connects seamlessly with the existing networking protocols and cost effectively enables connectivity solutions among CE peripherals will shift the home entertainment environment. It will enable multiple usage models from cable replacement to the streaming of video, audio, and other entertainment media [3]. Many UWB components and systems are already in the testing and demonstration phases, with actual release dates for final consumer products expected in early 2005. Many Corporations are working with the industry to enable this exciting technology and help ensure its success.

REFERENCES

[1] Ultra-Wideband Technology and Test Solutions Technical Brief, Tektronix, [http://www.afcingenieros.com/uploads/Afc/InfoTecn/RTSA/pdf/14%](http://www.afcingenieros.com/uploads/Afc/InfoTecn/RTSA/pdf/14%20Ultrawideband%20Technology%20and%20Test%20Solutions%20-%20AFC.pdf)

20Ultrawideband%20Technology%20and%20Test%20Solutions%20-%20AFC.pdf, Accessed online on 15-09-2010.

[2] Yasir Faheem, Saadi Boudjit, Ken Chen, Wireless Multimedia Sensor Networks: Requirements & Current Paradigm, <http://www.portailthd.fr/media/uploads/2010/06/24/fiche-reseaux-de-capteurs.pdf>, March 8, 2010

[3] Mrutyunjaya Panda and Manas Ranjan Patra, Simulation study of Ultra Wideband Modulation system, in Proc. of International conference on Communication and Power systems (ICCPs), Chennai, Dec. 14-16, 2006, pp. 158-162.

[4] Ping Liu and Zhengyuan Xu, POR-Based Channel Estimation for UWB Communications, IEEE Transactions on Wireless Communications, Vol. 4, No. 6, November 2005, pp. 2968-2981.

[5] K. L. Yong, B. E. Toh, F. N. Lai, and G. L. Tan, "Coexistence study on ultra-wideband and mobile cellular systems," in Proc. IEEE Int. Conf. on Ultra-Wideband (ICU), Zurich, SWITZERLAND, Sept. 2005, pp. 752-757.

[6] Q. Li and L. A. Rusch, "Multiuser Receivers for DS-CDMA UWB," Proc. IEEE 2002 Conf. on UWB Syst. and Technology, Baltimore.

[7] C. Corral, et al., "Pulse Spectrum Optimization for Ultra-Wideband Communication," Proc. 2002 IEEE Conf. On Ultra Wideband Systems and Technologies (UWBST 2002), 21-23 May 2002, Baltimore.

[8] R. Fontana et al., "Recent Advances in Ultra Wideband Communications Systems," Proc. 2002 IEEE Conf. On Ultra Wideband Systems and Technologies (UWBST 2002), 21-23 May 2002, Baltimore.

[9] R. J. Fontana and S. J. Gunderson, "Ultra-Wideband Precision Asset Location System," Proc. IEEE 2002 Conf. on UWB Syst. and Technology, Baltimore.

[10] D. Porcino and G. Shor, "Wireless Video Links via UWB Technology: Application Requirements," IEEE 802.115 document SGAP3a-02/119r0. Available online at: <http://grouper.ieee.org/groups/802/15/pub/2002/Mar02/>.

[11] D. Cassioli et al., "Performance of Low-Complexity Rake Reception in a Realistic UWB Channel," Proc. IEEE 2002 International Comm. Conf., pp. 763-767.

[12] F. Ramirez-Mireles, "On Performance of Ultra Wideband Signals in Gaussian Noise and Dense Multipath," IEEE Trans. on Vehicular Technology, January 2001, pp. 244-249.

[13] "Propagation Measurement," web page of UltraLab at the University of Southern California: http://ultra.usc.edu/New_Site/propagations.html

[14] J. D. Taylor, "Ultra-Wideband Radar Overview," in Introduction to Ultra-Wideband Radar Systems, CRC Press.

[15] M. Welborn, "Multiple Access Options for UWB WPANs," presentation to IEEE 802.15 high-rate WPAN alternate PHY study group (SG3a), document 02/382. Available online at <http://grouper.ieee.org/groups/802/15/pub/2002/Sep02/>

[16] L. B. Michael, M. Ghavami, and R. Kohno, "Multiple Pulse Generator for Ultra-Wideband Communication using Hermite Polynomial Based Orthogonal Pulses," Proc. 2002 IEEE Conf. On Ultra Wideband Systems and Technologies (UWBST 2002), 21-23 May 2002, Baltimore

[17] M. Z. Win and R. A. Scholtz, "Impulse Radio: How It Works," IEEE Comm. Letters, vol. 2, pp. 10-12 (January 1998).

[18] Revision of Part 15 of the Commission's rules regarding UWB transmission system, [http://www.fec.gov/bureau/engineering technology/orders/2002/fcc02048.pdf](http://www.fec.gov/bureau/engineering%20technology/orders/2002/fcc02048.pdf)

Performance of botanical pesticides to control post-harvest fungi in citrus

H. Singh, *G.Alsamarai, M. Syarhabil
School of Bioprocess Engineering
University Malaysia Perlis

Contact e-mail:ghassan_alsamarai@yahoo.com

Abstract— Three alcoholic extracts from *Cerbera odollam* L. (Suicide tree), *Syzygium aromaticum* L. (Clove) and *Swietenia macrophyllai* L. (Mahogany) at concentrations of 500 ppm, 1000 ppm, 2000ppm and 3000ppm, were tested for antifungal activity *in vitro* on *Penicillium digitatum*, *Aspergillus niger* and *Fusarium sp* isolated from naturally infected citrus fruit. The water extracts served as control. Results show the alcoholic extract concentrations were more effective than the water extract control in showing antifungal activity ($P < 0.05$) against test pathogens. All 3000ppm concentration from *Cerbera odollam* L. showed a 90% inhibition zone for all the three fungi. The inhibition zone of *Syzygium aromaticum* L. and *Swietenia macrophyllai* L. were 40% and 60% respectively, at the same 3000 ppm concentration. Plant extracts are viable alternatives to chemical pesticides; they are readily available non-pollutive, cost effective, non-hazardous, and they do not disturb ecological balance. Moreover, Investigation are to test the efficacy of these extracts practical application.

Index Terms— Plant extracts, *Cerbera odollam* L., Post-harvest pathogens, Disease management, Botanical biopesticide

1 INTRODUCTION

Post-harvest diseases account for 50% of losses in fruits stored under poor storage conditions especially under high humidity. They pose a major problem to the agriculture industry (Agrios, 2005). Citrus fruits are one of the crops susceptible to post-harvest diseases caused by fungi under poor storage conditions. The most important fungi causing post-harvest diseases include: *Penicillium spp*, *Aspergillus niger*, *Monilinia lax*, and *Rhizopus stolonifer* (Ogawa *et al.*, 1995). Many fruits are prone to damage caused by insects, animals, early splits, and mechanical harvesting. The damage predispose the fruit to wound invading pathogen *Aspergillus flavus*, and other fungi, that causes the decay to spread in stored citrus fruits. *Aspergillus flavus* can pose as a serious health problem because of its production of aflatoxin, which is a group of toxic and carcinogenic compounds (Diener *et al.*, 1987; Wilson and Payne 1994; Palumbo *et al.*, 2006).

Synthetic fungicides, such as, thiabendazole, imazalil and sodium ortho-phenyl phonate (Poppe *et al.*, 2003) have been used traditionally to control postharvest diseases. However, their excessive use, high cost, residues in plants, and development of resistance, have left a negative effects on human health and the environment (Paster and Bullerman, 1988; Bull *et al.*, 1997). Environmentally friendly plant extract agents have shown great potential as alternatives to synthetic fungicides (Janisiewicz and Korsten, 2002; Zhang *et al.*, 2005). Recently, the antimicrobial activity of biodegradable and safe higher plant products (Kumar *et al.*, 2008) has attracted the attention of microbiologists. However, the actual use of these products to control postharvest pathogens of fruits, particularly citrus pathogens, is still limited. The purpose of the current research is to test the possibility of using extracts from Suicide,

Clove, and Mahogany trees to control or inhibit post-harvest diseases causing pathogens in citrus fruits.

2. MATERIALS AND METHOD

2.1 COLLECTION OF DISEASED FRUITS

Wet markets at Kangar (Perlis) and Georgetown (Penang) were surveyed in December 2010, to observe common post-harvest disease symptoms in oranges, lemons, and grape fruits. The prominent symptoms observed were the growth of green, black, white, or blue colored molds on the fruits. Random samples were collected from citrus fruits and brought to the Microbiology laboratory of the School of Bioprocess Engineering, University Malaysia Perlis for further studies. The fruits were washed with water, disinfected with 10 % sodium hypochlorite, and cultured in sterilized PDA media under aseptic lamina conditions, for identification, single-spore isolation, and propagation under laboratory conditions at 25°C.

2.2 PATHOGENS

The pathogens identified using the taxonomic and morphological references were *Aspergillus niger*, *Penicillium digitatum*, and *Fusarium sp*. Highly aggressive, single-spore isolates of *P. digitatum*, *A.niger* and *Fusarium sp*. originally isolated from citrus fruits were grown on potato dextrose agar (PDA) at 25°C for 7 days. The spores were harvested by flooding the media surface with distilled water and gently agitating the plate to dislodge spores (Obagwu and Korsten, 2002). The spores were then refrigerated for further studies and propagation.

2.3 PREPARATION OF PLANTS FOR EXTRACTION

Cerbera odollam L (Suicide tree), *Syzygium aromaticum* L. (Clove), and *Swietenia macrophyllai* (Mahogany) were col-

lected from a kitchen garden housing-estate Kangar. The collected samples were washed under running water, to get rid of dirt, insects and plankton. Subsequently they were dried overnight in the laboratory-electric oven at 40c. One 100g of the material (leaves and fruits) were pulverized using an electric mixer, and preserved in labelled glass which were sealed until use.

2.4 PREPARATION OF PLANT EXTRACTS

The extraction technique used was a modification of Ruch's (2001) method. Up to 50g each of the oven dried and pulverized powered material from *Azadirachta indica* (Neem), *Cymbopogon citratus* (Lemon grass) *Zingiber officinale* (Ginger), *Cap-sicum frutescent* (Chilly) *Syzygium aromaticum* (clove) were treated with 500 ml of 95% alcohol with constant stirring for 30 min. After stirring, the solutions were filtered through 2 layers of cheese-cloth gauze and Whitman's (No.2) filter paper before the filtrates were subjected to evaporation through Rotary Evaporator at 60°C degree for 60 min. The dark spongy materials from the Rotary evaporator were removed and dried in an oven at 37°C for 2 days. The dried powder was stored in small and sterilized 5ml screw-capped glass bottles they were refrigerator (4°C) until further use.

2.5 PREPARATIONS OF PLANT EXTRACT DILUTIONS

The Suicide tree, Clove and Mahogany powder extracts were removed from the refrigerator and were brought to the lab for the preparation of extract dilutions. Aliquots of 1.0g, 2.0g an3.0g from each powder (plants) were mixed with organic solvent dim ethyl sulfoxide (DMSO) to obtain the concentrations required after the complete volume with distilled water to make dilutions of 500 ppm, 1000 ppm, 2000 ppm, and 3000 ppm.

2.6 IN VITRO SCREENING

PDA media was incorporated in forty-five 50 ml glass flasks and autoclaved for 20 min. After autoclaving, the flasks were cooled to about 45°C. Approximately 5ml of plant extract,(500 ppm, 1000 ppm, 2000 ppm, and 3000 ppm) were taken form the Suicide tree, Clove, and Mahogany. They extracts were pipetted into four of the forty-five 50 ml flasks and were gently agitated by hand for 2 min for a proper mixing of extract. Up to 20 ml aliquots of the mixed media were dispensed into 9cm petri-dishes. Subsequently chloramphenicol (250 ml/g per petri dish) was added to the medium to prevent bacterial growth (Nikos *et al.*, 2007). The experiment was performed under aseptic lamina conditions and replicated thrice. Approximately 1ml from *P. digitatum*, *A.niger* and *Fusarium. Sp* spore suspensions (conc.1 × 10⁶ spores/ml) were pipetted on the center of the amended PDA extracts. The inoculated plate was then incubated at 25°C for 10 days. The petri-dishes inoculated without the extract concentrations, served as control. Moreover colony diameter was determined by measuring the average radial growth. The inhibition zone (P), was measured using the formula of Francisco (2010):

$$P = \frac{C - T}{C} \times 100$$

Where C is the colony cm² of the control and T is of the treatments (three replicates).

3. STATISTICAL ANALYSIS

The experimental data was subjected to analysis of variance (ANOVA). Significant differences between mean values were determined using Duncan's Multiple Range test (P= 0.05) following ANOVA. Statistical analyses were performed using SPSS (SPSS Inc., Chicago, USA).

4. RESULT

The post-harvest fungi which were identified on the basis of their cultural and morphological characteristics, and tested for anti microbial activity were *Penicillium digitatum*, *Aspergillus niger*, and *Fusarium sp.* Mixing culture PDA media with all concentrations (0 ppm (control), 500 ppm, 1000 ppm, 2000 ppm, and 3000 ppm) of the plant extracts from *Cerbera odollam* L. showed significant results (P>0.05, Fig. 1) compared with the control. *Penicillium digitatum* showed a reduction in colony development in ascending order; ranging from 69.3%, 77.8 %, 83.7 %, and 93 % at concentrations of 500, 1000, 2000, and 3000 ppm respectively. *Aspergillus niger* recorded inhibition zones of 74%, 80.7%, 86.6%, and 95.1% at similar plant extract concentrations. The inhibition zones observed in *Fusarium sp* were 57.7% 68.5%, 73.1% and 95%. No inhibition zone was seen in control treatments. The results show that 3000 ppm achieved the best results in inhibiting mycelial growth among three fungi the three fungi studied.

Result of the efficacy of plant extracts on post-harvest pathogens in citrus fruits are presented in Figures.1, 2, and 3. A different trend in the microbial inhibition activity (P>0.05) of the *Cerbera odollam* L. extract was observed with all three fungi (*Aspergillus niger*, *Penicillium digitatum*, and *Fusarium sp*) except at 3000ppm.

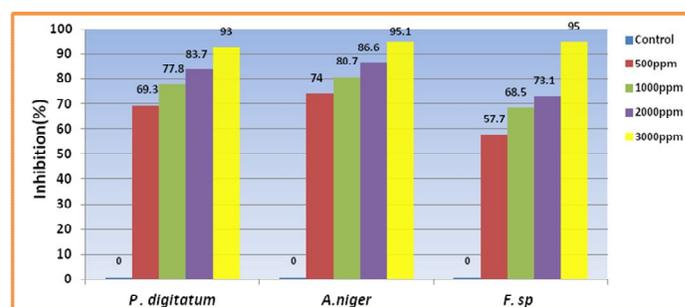


Figure1. Impacts of ethanolic extract of *Cerbera odollam* L. (Suicide tree) expressed as % of inhibition zone on colony growth (cm²) of *Penicillium digitatum*, *Aspergillus niger* and *Fusarium sp*, raised on PDA and incubated at 25°C.

Tret (ppm)	<i>C. odollam</i>			<i>S.aromaticum</i>			<i>S.macrophyllai</i>		
	<i>P.d</i>	<i>A.n</i>	<i>F.sp</i>	<i>P.d</i>	<i>A.n</i>	<i>F.sp</i>	<i>P.d</i>	<i>A.n</i>	<i>F.sp</i>
cont	9.00	9.36	6.55	8.13	9.26	6.66	9.00	9.46	6.56
500	2.76	2.70	2.77	5.96	5.90	5.16	4.40	4.66	4.33
1000	1.99	1.80	2.06	5.74	5.56	4.46	4.06	3.93	2.56
2000	1.46	1.26	1.76	5.13	5.13	4.46	3.20	3.56	3.26
3000	0.63	0.45	0.22	4.76	4.86	4.03	2.96	2.90	2.48

control treatment. The impact medial of *Swietenia macrophyllai* L. in the inhibition zone on colony growth (60%) is also given

Table 1. Impacts of extracts of *Cerbera odollam* L., *Syzygium aromaticum* L. and *Swietenia macrophyllai* plant extracts on Colony Diameter [CD] in cm of *Penicillium digitatum*, *Aspergillus niger* and *Fusarium sprose* on PDA

CD* refers to colony diameter *P.d*= *Penicillium digitatum*
A.n= *Aspergillus niger* *F.sp* = *Aspergillus niger*

5- Discussion

The objective of the current research is to study the effect of plant extracts on the mycelia growth of, *Penicillium digitatum*, *Aspergillus niger* and *Fusarium sp* that are pathogens of the post-harvest diseases of citrus fruits, as reported by Eckert & Sommer,(1967),and Adaskaveg *et al*, (2002). These diseases have caused 10-30% decrease in crop yield and marketing quality (Agrios, 2005, and Serrano *et al.*, 2005).

The use of biocontrol agents from plant extracts like lemon, citronella, clove, mint, thyme and oregano oils has been employed by Viudamartos *et al* (2007), as alternatives for conventional synthetic pesticides in plant disease control. The plant extracts that are reportedly effective against the fungi *Penicillium digitatum* include garlic (Obagwa,2002), neem (Mossini, *et al.*,2009), *Withania somnifera* and *Acacia seyal* (Samson, 1984),and mustard and horseradish (McOnie,1964).

Clove completely inhibits the mycelia growth of *A. flavus* and aflatoxin formation (Karapynar, 1989) . *Aspergillus niger* has been noted for its carcinogenic aflatoxin production in diseased plants. Montes and Carvjal (1998) , in their research for involves screening more than 280 plant species for their inhibitory effect on the toxin, have reported that about 100 of these plants have some activity on the growth of toxin production by fungi

Garlic extract has a positive effect on *Fusarium* inhibition (Anjorin *et al.*, 2008). Saxena and Mathela (1996), in their study on the inhibitory ;

effect of plant extracts on *Fusarium*, have reported that, *Azadirachta indica*, *Artemisia annua*, *Eucalyptus globules*, *O.cimum*, *Sanctum* and *Rheum emod* , have shown significant reduction of pathogens. In the current research, the Suicide tree at 3000 ppm has been discovered to shows almost 95 % inhibition of the mycelia growth in culture medium.

CONCLUSION

Most plant derivatives, phenols and alkaloids tend to show positive effect on the inhibition of postharvest fungal or bacterial pathogens. Amidst an increasing global environmental pollution, these plant extracts or botanicals have great replace potential replacing conventional synthetic pesticides in the future.

ACKNOWLEDGEMENTS

The researchers wish to thank the Scholl Bioprocess Engineering, Universiti Malaysia Perlis for providing facilities to conduct this research.

REREFRANCES

[1] Adaskaveg, J.E., H. Forster, N.F. Sommer. (2002). Principles of post-harvest pathology and management of decays of edible horticultural crops. In: *Post-harvest Technology of Horticultural Crops*, (Eds.): A.Aader. Vol. 3311. Universi-

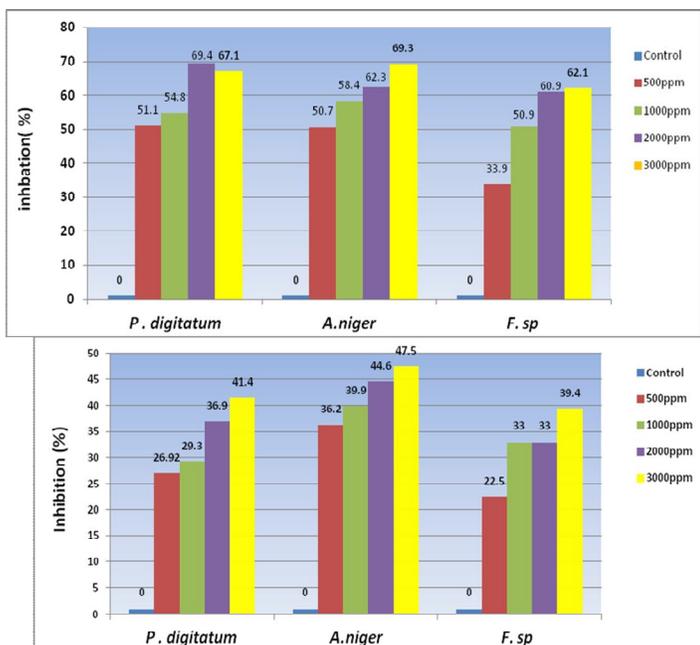


Figure 2. Impacts of ethanolic extract of *Syzygium aromaticum* L. (Clove) expressed as % of inhibition zone on colony growth (cm²) of *Penicillium digitatum*, *Aspergillus niger* and *Fusarium sp* raised on PDA and incubated at 25°C.

Figure 3. Impacts of ethanolic extract of *Swietenia macrophyllai* L. (Mahogany). expressed as % of inhibition zone on colony growth (cm²) of *Penicillium digitatum*, *Aspergillus niger* and *Fusarium sp* raised on PDA and incubated at 25°C.

To modify the running headings, select View Header and Footer. Click inside the text box to type the name of the journal the article is being submitted to and the manuscript identification number. Click the forward arrow in the pop-up tool bar to modify the header or footer on subsequent pages. The impact of concentrations on the inhibition diameters of the fungi are presented in Table 1. The data shows different effects. *Syzygium-aromaticum* L. gives slow inhibition zones (40%) for both extracts at concentrations of 500ppm to 3000 ppm, in contrast with the

- ty of California Publication, California, Pp. 163-195. Agrios, G.N., (2005). *Plant Pathology*, Academic Press, New York.
- [2] Anjorin, S.T., H.A. Makun, T. Adesina and I. Kudu.,(2008). Effects of *Fusarium verticillioides*, its metabolites and neem leaf extract on germination and vigour indices of maize (*Zea mays* L.) .*Afr. j.Biotechnol.* 7:2402-2406.
- [3] Bull, C.T, Stack, J.P, Smilanick, J.L.,(1997). *Pseudomonas syringae* strains ESC-10 and ESC-11 survive in wound on citrus and control green and blue molds of citrus. *Biol. Contr.* 8: 81-88.
- [4] Diener, U.L, Cole, R.J, Sanders, T.H, Payne, G.A, Lee, L.S, Klich, M.A.,(1987). Epidemiology of aflatoxin formation by *Aspergillus flavus*. *Ann. Rev. Phytopathology.* 25: 249-270.
- [5] Eckert, J.W. and N.F. Sommer.(1967)., Control of diseases of fruits and vegetables by post-harvest treatment. *Ann. Rev. Plant Pathol.*, 5: 391-432.
- [6] Francisco, D.H., (2010). *Lippia graveolens* and *Carya illinoensis* Organic Extracts and there in vitro Effect against *Rhizoctonia Solani* Kuhn. *American Journal of Agricultural and Biological Sciences*5(3): 380-384.
- [7] Janisiewicz, W.J., Korsten, L.,(2002). Biocontrol of post harvest diseases of fruits. *Annu. Rev. Phytopathology*40: 411-441.
- [8] Obagwa, J., Korsten, L., (2002). Control of citrus and green and blue molds with garlic extracts. *Plant Pathology*.109, 221- 225.
- [9] Karapynar.,(1989). Inhibition effects. Of some spice agents on aflatoxigenic mould growths. *Proceedings of the International Food Symposium, April 4-6, Bursa, Tukey, pp:129-137.*
- [10] Kumar, A., Shukla, R., Singh, P., Prasad, C.S., Dubey, N.K., (2008). Assessment of *Thymus vulgaris* essential oil as a safe botanical preservative against post harvest fungal infestation of food commodities. *Innu. Food Sci. Emerg.*, 4: 575-580.
- [11] Ogawa, J.M., Dehr, E.I., Bird, G.W., Ritchie, D.F., Kiyoto, V., Uyemoto, J.K., (1995). *Compendium of Stone fruit Diseases*. APS Press, USA.
- [12] McOnie, K. C., (1964). The latent occurrence in citrus and other hosts *Gignardia* easily confused with *G. citricarpa*, the black spot pathogen. *Phytopathology*54:40- 43.
- [13] Mossini, S.A.G., C. Carla and C. Kimmelmeier.,(2009). Effect of neem leaf extract and Neem oil on *Penicillium* growth, sporulation, morphology and ochratoxin A production. *Toxins*, 1: 3-13.
- [14] Nikos, G.T and C. D. Economakis.,(2007).Antifungal activity of Lemongrass (*Cymbopogon citrates* L.) essential oil against key postharvest pathogens. *Global Journal of Biotechnology & Biochemistry*3 (2): 56-59, 2008.
- [15] Palumbo, J.D., Baker, J.L., Mahoney, N.E., (2006). Isolation of bacterial antagonists of *Aspergillus flavus* almonds. *Microbial Ecol.* 52(1): 45- 52.
- [16] Paster, N., & Bullerman, L. B., (1988). Mould spoilage and mycotoxin formation in grains as controlled by physical means. *International Journal of Food Microbiology*, 7, 257–265.
- [17] Poppe, L., Vanhoutte, S., Höfte, M., (2003). Modes of action of *Pantoea agglomerans* CPA-2, an antagonist of postharvest pathogens on fruits. *Eur. J. Plant Pathol.* 109: 963-973.
- [18] Ruch, B.a. Worf, R.(2001) . Processing of neem for plant protection simple and sophisticated standardized extracts. Abstracts of the .Work shop, Neem and Pheromones , University of Uberaba ,Brazil, March 29-30 Augusts, P.499.
- [19] Samson, J.A., (1984).Tropical fruits- Tropical agricultural series. Longman Inc., New York, pp.64-118.
- [20] Saxena. and C.S. Mathela. (1996). An activity of new compounds from *Nepeta leak* and *Nepeta clarkei*. *Applied Environ Microbiol.* 702- 704.
- [21] Serrano, M., D. Martinez-Romero, S. Castillo. Guillen and D. Valero., (2005). The use of the natural antifungal compounds improves the beneficial effect of MAP in sweet cherry storage. *Innovative Food and Emerging Technologies*, 6: 115-123.
- [22] Wilson, D.M., Payne, G.A.,(1994). Factors affecting *Aspergillus flavus* group infection and aflatoxin contamination of crops, In the Toxicology Aflatoxins. Human Health, Veterinary, and Agricultural Significance (Eaton DL, Groopman JD, San Diego: Academic Press).
- [23] Viudamartos, M. Ruiz-navajas, Y., Fernandez -lopes, J. and Perez-alvarez, J.A.,(2007). Antifungal activities of thyme, clove and oregano essential oils. *Journal of Food Safety*27: 91-101.
- [24] Zhang, H., Zheng, X.,(2005). Biological control of postharvest blue mold of oranges by *Cryptococcus laurentii* (Kufferath) Skinner. *Biocontrol* 50:331-342.

Stabilization Assessment of Aggregates in Asphalt Concrete Mixtures

I. Akiije, G. L. Oyekan

ABSTRACT - Aggregates particle size and gradation assessment of asphalt concrete mixtures produced in the south-western part of Nigeria have been investigated for mechanical stabilization. Samples of asphalt concrete produced were collected from three selected asphalt plants (A, B, and C) all located in the south-western part of Nigeria. Bitumen Extraction and Aggregates Sieve Analysis were carried out on the collected samples. The result of the investigation has shown that from each of the asphalt plant, the maximum aggregate size used in the production of asphalt concrete was 12.5 mm while the nominal size was 9.5 mm. However, the asphalt concrete paving materials investigated were of a dense mixture because of the presence of mineral materials that retained on sieve openings 2.36 mm, 0.3 mm, 0.15 mm and 0.075 mm. Sample from asphalt plant A was with aggregates gradings approaching the maximum amount permitted to pass the 2.36 mm sieve openings and this will result in pavement surfaces having comparatively fine texture. While samples from asphalt plants B and C were with aggregates gradings approaching the minimum amount passing the 2.36 mm sieve openings which will result in surfaces with comparatively coarse texture.

KEYWORDS: Aggregates, Asphalt Concrete, Dense Mixture, Sieve Openings, Specifications



1 INTRODUCTION

Mechanical stabilization is the selection and combination of aggregates to obtain a gradation within the limits of any mix design of asphalt concrete. Gradation is the designed aggregate structure defined by the distribution of aggregate particle sizes for a given blend of aggregate mixture. Maximum size gradation is defined as one sieve larger than the nominal maximum size. Nominal maximum size gradation is one sieve larger than the first sieve that retains more than 10 percent of the aggregate. Maximum density gradation is obtained when the aggregate particles fit together in their densest form. Asphalt concrete is a composite material of a uniformly mixed combination of asphalt cement and aggregates. Asphalt concrete can be used in variety of ways, including the construction of highways and airport pavement and bases, parking areas, and industrial floors. Asphalt cements are solid hydrocarbons with certain physiochemical characteristics that make them cementing agents obtained after separation of the lubricating oils. They are also very viscous, and when used as a binder for aggregate in pavement construction, it is necessary to heat both the aggregate asphalt cement prior to mixing the two materials. Asphalt cements are used mainly in the manufacture of hot-mix, hot-laid asphalt concrete.

G. L. Oyekan. Dept. of Civil and Environmental Engineering, University of Lagos, e-mail: gloyekan@yahoo.com

Aggregates refer to granular mineral particles of a composite material that resists compressive stress and provides bulk to the composites. Aggregates can be sourced from natural deposits of sand and gravel, pulverized asphalt pavements, crushed stone and blast-furnace slag. Aggregates are widely used for highway base, subbase and backfill. According to Adedimila [1] aggregates are used in combination with asphalt or bitumen, a cementing material, to form base and binder courses as well as surface and levelling courses.

In Nigeria premature failure of highway pavement shortly or less than 5 years after construction or rehabilitation is a thing of concern. This is a major challenge to highway engineers, contractors and the government of Nigeria because designed life span of a road is 20 years. Substandard materials, poor design mix of highway pavement components and inadequate quality control among other factors are bane in meeting the standard stated in the specifications of roads. Ajani [2] indentified the causes of asphalt pavement failure to include; use of substandard road materials, lack of current traffic studies and analysis, insufficient hydrological data, insufficient geo-technical studies, use of inappropriate design, inadequate quality control, inadequate Supervisory/Technical Staff.

I. Akiije. Department of Civil and Environmental Engineering, University of Lagos, e-mail: akiijeia@yahoo.com

The aim of this study therefore is to determine the stability of the selected and combined aggregates used for a defined asphalt concrete whilst meeting the required standard specifications. Specifically the objectives of this study include (i) investigation into the aggregates particle size and gradation properties of some produced asphaltic concrete (ii) comparison of the results of (i) with a standard specifications (iii) Make inferences on the results and make appropriate recommendations, including areas of future research work. The scope of this research project includes the collection of asphaltic concrete samples from three different asphalt producing plants located in the south-western part of Nigeria. Also, bitumen extraction test on each of the asphaltic concrete material samples to check the sieve analysis of the constituents aggregates and checking with specified standards. The investigation reported by this paper is limited to aggregates particle size and gradation assessment of already prepared asphalt concrete mixtures after bitumen extraction. The significance of this study is that grain-size analysis results data plotted on an aggregate grading chart will serve as useful aid at determining the gradation of aggregates used for highway projects and conformance to the limits of a standard specification.

American Society for Materials and Testing Standard specification developed by an official body in the United States of America is considered here for processing and testing methodology of aggregate materials. Table 1 gives suggested grading requirements of aggregate material based on the ASTM Designation 3515 [3] and Kentucky Department of Highways [4].

Considering the total grading characteristics of a bituminous paving mixture, the amount passing the 2.36 mm sieve is a significant and convenient field control point between fine and coarse aggregate. Gradings approaching the maximum amount permitted to pass the 2.36 mm sieve will result in pavement surfaces having comparatively fine texture, while coarse gradings approaching the minimum amount passing the No.8 sieve will result in surfaces with comparatively coarse texture.

Plotting of the grain-size analysis data are usually done on an aggregate grading chart as demonstrated in Figure 1.

2 MATERIALS AND METHODOLOGY

Samples of asphalt concrete produced were collected from three (A, B, and C) selected asphalt plants all located in the south-western part of Nigeria. Bitumen Extraction and Grain-Size Analysis Experiments for gradation of aggregates were carried out on the collected asphalt concrete samples. The

The chart is a powerful aid to make it easy for engineers to determine a preferred aggregate gradation. The chart is necessary so that the gradation of aggregates for highway projects conforms to the limits of a standard specification band.

TABLE 1: AGGREGATE GRADING REQUIREMENTS FOR DENSE MIXTURE BITUMINOUS PAVEMENTS

Passing Sieve Designation	Percentage Passing, by weight	
	Surface Course Specification from ASTM Designation 3515*	Surface Course Specification from Kentucky Department of Highways**
19 mm	100	100
12.5 mm	90 to 100	80 to 100
9.5 mm	56 to 80	55 to 80
4.75 mm	44 to 74	35 to 60
2.36 mm	28 to 58	20 to 45
0.3 mm	5 to 21	3 to 14
0.15 mm	3 to 20	2 to 7
0.075 mm	2 to 10	

SOURCES: *Garber, N. J., and Hoel, L. A. [5]

** Wright, P. H. [6]

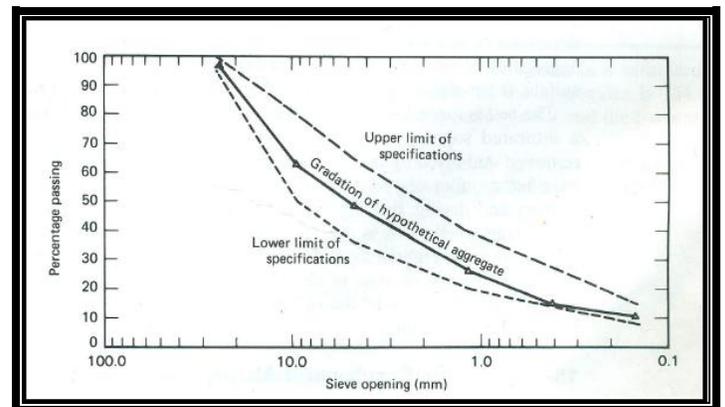


Figure 1: Aggregate Gradation Specification Chart

Sources: Wright, P. H. [6]

main objective of bitumen extraction test is to provide asphalt free aggregates for gradation analysis for this study. It is worthy of note that aggregates gradation is the blend of particle sizes in the mix which influence the density, strength, and economy of the pavement structure.

2.1 Bitumen Extraction Experiment

Bitumen extraction experiment was carried out for the purpose of separating the asphalt from the mineral aggregates in the asphaltic paving mixture obtained from the selected asphalt plants A, B, and C which are all located in the south-western part of Nigeria.

From each of the asphalt concrete sample, 1000 g of was weighed and bitumen extraction test was carried out respectively in triplicate by order of Sample A, Sample B and Sample C. The apparatus used include bitumen extracting machine, weigh balance, trowel, weighing-pan, kerosene, stove and oven capable of maintaining a temperature of 110±5°C. Carbon disulfide and trichloroethylene solvent cannot be used because they may be life-threatening for they affect the nervous system, kidney, and they may cause liver damage, headaches, lung irritation, dizziness, poor coordination, and difficulty in concentrating. Also, skin contact with trichloroethylene for short periods may cause skin rashes.

In the bitumen extraction test carried out, weight of asphalt concrete (asphalt + aggregates), X, is 1000 g; weight of cleanup aggregates Y and weight of bitumen Z and therefore the percentage by weight of asphalt P% are calculated thus

$$X - Y = Z \tag{1}$$

$$(Z/X) \times 100 = P\% \tag{2}$$

2.2 Grain - Size Analysis Experiment of Extracted Combined Aggregates

Grain-size analysis to determine the relative proportions of various particle sizes in the mineral aggregate mix after bitumen extraction experiment was carried out. Grain-size analysis was used to determine the relative proportions of various particle sizes in the cleanup mineral aggregates mix for samples A, B and C. To perform this analysis, the weighed cleanup aggregates sample having dried up for 24 hours was shaken over a nest of sieves having selected sizes of square openings. The sieve openings used were 19 mm, 12.5 mm, 9.5 mm, 4.75 mm, 2.36 mm, 0.3 mm, 0.15 mm and 0.075 mm. The sieves were grouped together with the one with the largest openings on top and those with successively smaller openings placed underneath. The aggregate sample was shaken with a mechanical sieve shaker, and materials retained on each sieve opening with the exception of sieve opening 19 mm were weighed. The weight of material retained on each sieve was determined and expressed as a percentage of the original sample.

3.0 RESULTS AND DISCUSSION

The test results of the bitumen extraction experiment is in Table 2 of which three samples from asphalt plants A, B and C are having bitumen content on the average of 5.4 %, 5.8 % and 6.2 % respectively. The results of gradation specification analysis of the extracted combined aggregates of asphalt samples from plants A, B, and C are in Table 3. Graphically, the results of the gradation specification analysis of the extracted combined aggregates of asphalt samples from plants A, B, and C are also plotted on an aggregate grading chart as shown by Figure 2.

TABLE 2: SUMMARY OF THE PROPERTIES OF THE ASPHALT CONCRETE SAMPLES IN COMPARISON WITH THE SPECIFICATIONS

Properties	Specification From ASTM Designation 3515	Sample A	Sample B	Sample C	Remarks
Bitumen content by weight % of total mixture	4 to 11	5.4	5.8	6.2	Each sample met the specification requirements

The results of respective bitumen extraction experiment of samples A, B and C are meeting the amount of 4 to 11 as standard values of the specifications from ASTM designation 3515 as shown in Table 2.

Considering Table 1, it is worthy of note that surface course aggregates specification from ASTM Designation 3515 used for samples A, B, and C made a useful comparison rather in this study rather than specification from Kentucky Department of Highways. This is vividly shown in Table 3. The reason is that ASTM Designation 3515 is having specification for percentage passing by weight of aggregates for sieve designation 0.075 by taking care of the available extracted aggregates of asphalt concrete paving stones samples. Whereas, specification on percentage passing of aggregates by weight for asphalt surface course, from Kentucky Department of Highways is lacking sieve openings with designation 0.075 that is very useful in this experiment. Also, it is obvious that the gradation specification analysis of the extracted combined aggregates of asphalt samples from asphalt plants A, B, and C did not conform to standard specification in particular ASTM designation 3515 as expressed in Table 3. In addition, extracted combined aggregate samples A, B, and C of Figure 2 are not in conformity with aggregates gradation specification chart of Figure 1. This is because in Figure 2 the result lines showing the gradation of samples A, B and C are not well defined

mid-way between the Lower and Upper limits of standard specifications as recommended in Figure 1

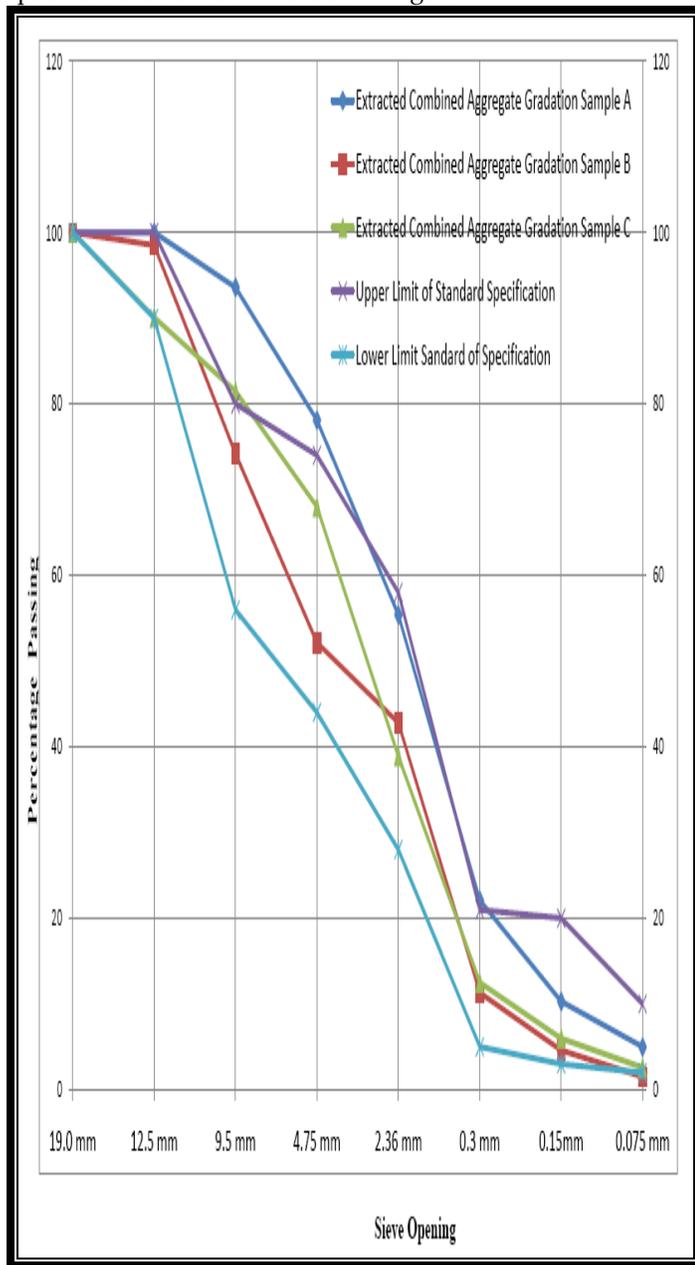


TABLE 3: GRADATION SPECIFICATION ANALYSIS OF THE EXTRACTED COMBINED AGGREGATES OF ASPHALT SAMPLES FROM PLANTS A, B, AND C

Passing Sieve Designation	Specification from ASTM Designation 3515	Sample A Percent By Weight	Sample B Percent By Weight	Sample C Percent By Weight
19 mm	100	100	100	100
12.5 mm	90 to 100	100	98.50	90.10
9.5 mm	56 to 80	93.60	74.20	81.50
4.75 mm	44 to 74	78.10	52.14	68.00
2.36 mm	28 to 58	55.40	42.80	38.90
0.3 mm	5 to 21	22.10	11.30	12.50
0.15 mm	3 to 20	10.30	4.60	6.00
0.075 mm	2 to 10	5	1.50	2.60

With the aid of the chart as in Figure 1, it is possible for an engineer to determine a preferred aggregate gradation for specified asphalt concrete at a glance. This is achievable in this study as in Figure 2 with the aid of Microsoft Excel where comparison of the gradation of combined aggregates extracted from already prepared asphalt concrete and the conformity to the band limits of Specification from ASTM Designation 3515 was possible easily.

Figure 2: Gradation Specification Analysis Chart of the Extracted Combined Aggregates of Asphalt Samples from Plants A, B, and C

4 CONCLUSIONS AND RECOMMENDATIONS

Based on the investigation carried out in this study the following conclusions and recommendations are derived as follows:

4.1 CONCLUSIONS

1. The determination of the stabilization assessment of aggregates gradation in asphalt concrete mixtures is better done by the extraction of asphalt from samples of asphalt concrete mixture for road pavement construction.
2. Better assessment aggregates gradation could be done by grain - size analysis experiment on extracted combined aggregates samples.
3. ASTM Designation 3515 Standard Specification methodology of aggregate gradation analysis has the desired ability of determining percentage passing stabilization assessment of aggregates gradation of asphalt concrete samples.
4. The three asphalt concrete paving materials investigated were of a dense mixture because of the presence of mineral materials that retained on sieve openings 2.36 mm, 0.3 mm, 0.15 mm and 0.075 mm.
5. Sample from asphalt plant A was with aggregates gradings approaching the maximum amount permitted to pass the 2.36 mm sieve openings and this will result in pavement surfaces having comparatively fine texture.
6. Samples from asphalt plants B and C were with aggregates gradings approaching the minimum amount passing the 2.36 mm sieve openings which will result in surfaces with comparatively coarse texture.
7. Supervising agencies of road pavement construction should learn from this study by ensuring the testing of the assessment of extracted aggregate stabilization of the asphalt concrete supplied for road pavement.
8. Varying cost of asphalt concrete supplied to pavement construction site could be easily determined by recognising the usefulness of this experimental investigation carried out in this study.
9. This study surely serves as prodigy towards the determination of strength and stability of asphalt concrete paving stones.

4.2 RECOMMENDATIONS

1. Stabilization assessment of extracted aggregates in asphalt concrete mixtures as carried out in this study must be done on samples of asphalt concrete from an asphalt plant yard before purchase and during the laying of asphalt concrete for road pavement.
2. The gradation of sample aggregate for standard stabilization assessment of extracted aggregates should have a graphic line result lying mid-way between upper and lower limits of standard specifications similarly as in Figure 1 and not necessarily like ones in Figure 2 so that the asphalt concrete paving stones produced is of good strength, durable and cost effective.

REFERENCES

- [1] A. S. Adedimila, "Bitumen: Nigeria's Other Black Gold?" Inaugural Lecture Series, University of Lagos, pp. 8, 2000.
- [2] A. R. Ajani, "Causes of Premature Failures on Nigeria Highways," Trans-Saharan Road Liaison Committee, Stage in Tunis, 2006.
- [3] ASTM D3515 "Hot Mixed, Hot-Laid Bituminous Paving Mixtures." Annual Book of American Society for Testing and Materials Standards, Section 4, Construction, Vol. 04.03, Road and Paving Materials; Pavement Management Technologies, American Society for Testing and Materials, Philadelphia, PA, 2007.
- [4] Kentucky Department of Highways, "Kentucky Standard Specifications for Road and Bridge Construction." Transportation Cabinet, Department of Highways, Frankfurt, KY, 1994.
- [5] N. J. Garber and L. A. Hoel, "Traffic and Highway Engineering." Fourth Edition Brooks/Cole Pacific Grove, 2010.
- [6] P. H. Wright, "Highway Engineering." Seventh Edition, John Wiley & Sons, New York, 2010.

An Improved Approach for Spatial Domain Lossless Image Data Compression Method by Reducing Overhead Bits

Mahmud Hasan, Kamruddin Md. Nur

Abstract— Lossless image compression techniques are used in digital imaging where large amount of data is to be stored without compromising the image quality. The volume of data that can be compressed using lossless image compression schemes is usually much lesser than that of its lossy compression counterparts. Yet, however, lossless compression algorithms are popular in a number of particular image data storage sectors. To meet the increasing demand of large amount of high quality image data storing, numerous algorithms were developed during last few decades featuring lossless image compression and covering various aspects of data compression approaches. Spatial domain lossless image compression methods are popular in most respects since their computational time is comparatively much lesser. In this paper, we focus on a spatial domain image compression technique that uses simple arithmetic operations in order to achieve the specified goal. We revealed that the mentioned algorithm is not always as advantageous as other spatial domain compression systems and often suffers from overhead transmission of unnecessary image data. The thorough investigation over the technique is reported along with the discovered mathematical bound at which the algorithm of interest is failed to achieve the desired target. Finally, to reduce the overhead obtained as a result of algorithmic trouble, an improved mechanism is suggested so that both the transmission time and storage space requirements using this method is facilitated.

Index Terms— Bits Per Pixel (BPP), Block Matrix, Block Processing, Computational Overhead, Inter-Pixel Redundancy, Run Length Coding, Spatial Domain Lossless Image Compression.

1 INTRODUCTION

RECENT digital imaging applications have observed copious invention in the field of image compression as the importance of preserving image data is being important day by day. Diverse ideas regarding this issue have been developed, still we suffer from choosing a suitable compression method for industry applications as computational cost of the compression stuff matters. A few standards like JPEG and JPEG-2000 are being used in today's industry applications where achieved compression ratio is important than its relative computational cost [1]. However, due to quality-compression trade-off, these standards fail to provide users with the most desirable image compression criterion- higher compression ratio with higher quality assurance [2].

Lossless image compression techniques, on the other hand, provides us with mentionable compression ratio and unaffected image quality. Such compression methods, that use simple arithmetic calculations in geometric domain or spatial domain [3,4,5], reduce the computational complexity too in a notable extent. Thus, spatial domain lossless image compression techniques deserve acute significance in digital imaging world. All of such algorithms attempt to reduce the inter-pixel redundancy of an image discovering the fact- since the value of any given pixel can be reasonably predicted from the value

of its neighbors, the information carried by individual pixels is relatively small [6].

In this paper, our prime focus is on a spatial domain lossless compression algorithm by Syed & Mehdi [7]. This algorithm performs lossless compression, although, often, it has to suffer from a large unnecessary amount of bits due to computational overhead. We investigated the reason and suggested necessary modification required to improve the algorithm. Comparative results have also been taken into account.

2 RELATED STUDY

In digital image compression terminologies, overhead bits refer to the extra amount of bits required by a specific algorithm to compress an image [8]. For example, let us suppose, we have an image compression algorithm that can reduce 500 bits from an image of 1000 bits. It is then regarded that the compression ratio achieved by this algorithm is $(1000 \div 500)$ for this particular image. Again, using the same algorithm, if another image of 800 bits results in 1000 bits after compression, there presents 200 overhead bits. It is possible since a good number of compression algorithms keep some non-image-information about the image in order to reduce its Bits Per Pixel (bpp). Whenever this non-image-information along with compressed-image-information becomes larger than original-image-information, an overhead occurs. Consequently, an image of n bits needs to be represented by m bits after compression; where $m > n$.

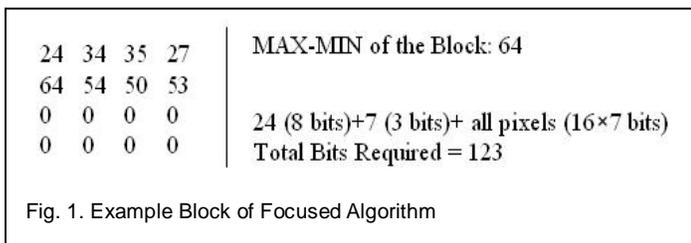
Preserving non-image-information in order to reduce total

-
- Mahmud Hasan is currently teaching at the Dept. of Computer Science & Engineering as a full-time lecturer in Stamford University Bangladesh. E-mail: hasanpoet@gmail.com
 - Kamruddin Md. Nur is currently teaching at the Dept. of Computer Science & Engineering as Assistant Professor in Stamford University Bangladesh. E-mail: kamruddin.nur@gmail.com

original-image-information is an age-old technique [9]. A commonly known algorithm for image data compression is Run Length Encoding (RLE) where a stream of same gray level pixels are encoded as (x,y) representing x as pixel-run and y as gray value [2,6]. In this case, x is not an original-image-information, rather, it is non-image-information to finally achieve compressed-image-information. A number of image compression algorithms are developed using such concepts as in [10, 11, 12]. There is another compression algorithm school that does not directly use RLE and preserve non-image-information. Rather, it preserves some non-image-information regarding a local $m \times n$ block so that this information helps decide the exact pixel value during decoding [7]. The algorithm we are going to investigate is categorized into this class of algorithms.

3 FOCUSED ALGORITHM

The algorithm we are examining has been developed by Syed & Mehdi [7]. It requires an image to be divided into a number of $m \times n$ blocks where the standard value of m and n is 4. The specific application has freedom to choose m and n other than 4. The algorithm then looks for the maximum and minimum pixel values MAX and MIN within this $m \times n$ block and calculates MAX-MIN. The block header preserves 8-bits MIN and a 3-bits coding-information that tells how many bits are required to represent MAX-MIN. Then MIN is subtracted from all pixel values of $m \times n$ block and each is encoded by k bits, where k is the number denoted by 3-bits coding-information. Figure 1 illustrates the focused algorithm. The embedding and extraction procedure as given by Syed & Mehdi [7] is shown in the following subsections.



3.1 Encoding Steps

The encoding steps of the algorithm proposed by Syed & Mehdi [7] are shown below –

- Step1: Select m and n for whole image.
- Step2: Take $m \times n$ non-overlapping block of image.
- Step3: Find the difference of Min and Max value in selected $m \times n$ block in X.

Step4: Add 11 bits header (8 bits for Min value of block, and 3 bits dedicated the no. of bits required to represent X value' in Y bits).

Step5: Subtract each pixel from Min value of a block and store in separate Y bits of every pixel in new $m \times n$ block.

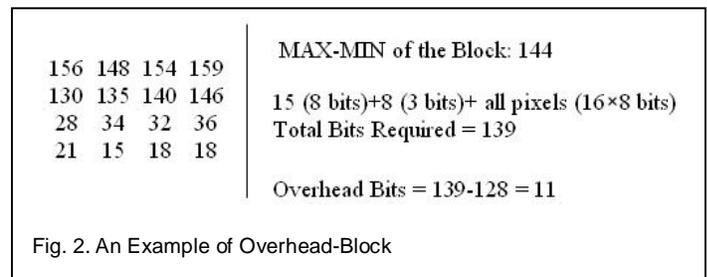
3.2 Decoding Steps

The decoding steps of the algorithm proposed by Syed & Mehdi [7] are shown below –

- Step1: Parse the header and find out block size m and n.
- Step2: Find the Min (8 bit) value for each block.
- Step3: Parse another 3 bits which represent the no. of (Y) bits required for each pixel value.
- Step4: Read next Y bits, add its value to Min and regenerate the actual value of pixel. Repeat this step for all pixels in a block.
- Step5: Repeat the above steps for whole image and regenerate the original image.

4 OVERHEAD ANALYSIS

It should now be clear that according to our focused algorithm any $m \times n$ block of an image contains a header of 11 bits, where MIN consists of 8 bits and coding-information comprises the left 3 bits. If these 3 bits denote 100_2 (i.e. 4_{10}), then every pixel of this block is encoded using 4 bits. This technique works efficiently as long as the coding-information remains less than 8 bits. But let us consider a situation where MAX-MIN results in an integer to represent which at least 8 bits are required. Then a typical 4×4 block has to be embedded as 11 bits+16*8 bits, whereas the non-compressed-block was embedded by only 16*8 bits. This situation is possible whenever $MAX-MIN \geq 128$. Figure 2 shows a practical phenomenon where such occurrence is illustrated.

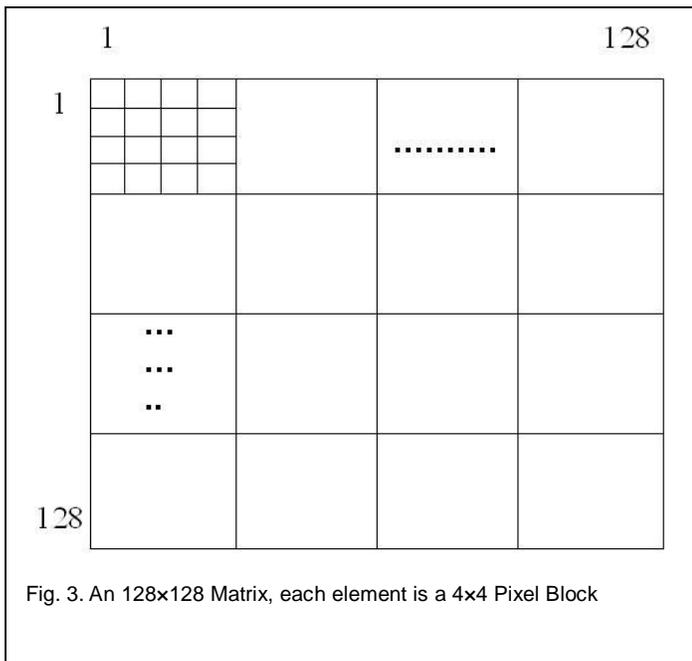


Although the spatial smoothness of an image is common, it is not guaranteed that at least two pixels of an $m \times n$ block cannot differ by a factor of 128 or more. Rather, it happens very frequently. Statistical evidence shows that for gray scale

images, there are at least 5.76% 4x4 blocks where $MAX-MIN \geq 128$. For color images, however, the percentage is less only 2.10%. Whatever the statistical percentage of overhead blocks, surely, for each overhead block, the focused algorithm needs to preserve some extra bits.

5 IMPROVED SUGGESTION

With a view to modifying our focused algorithm, we propose to keep some information regarding the overhead blocks and leave those blocks without embedding. Considering a standard block size 4x4 for 512x512 dimensional images, we find 16,384 blocks that can be treated as a 128x128 dimensional matrix as shown in Figure 3.



We require each row of this 128x128 matrix now starts with a 128 bit binary sequence where each 1 indicates an overhead block at that position. Since the overhead blocks are not embedded, 11 bits from each overhead block can be discarded. The CODEC ought now to use a trace variable that will keep checking the 128 bit block-row-header and whenever it finds a 1 in that row-header, it supposes no 11-bits block header for block of that position. For example, if 70th bit of 128 bits block-row-header contains a 1, for 70th block, the decoder does not look for 11-bits block header.

The improved encoding steps are organized as follows –

Step1: Prepend a 128 bit extra header in front of each block-row, all bits are reset.

Step2: Take a mxn non-overlapping block of image as done

in focused algorithm (standard size of m and n is 4).

Step3: Find the difference of Min and Max value in selected mxn block in X.

Step4: If $Max-Min \geq 128$ i.e. overhead block, set the corresponding bit in 128 bit header. Keep no 11 bit block-header.

Step5: Subtract each pixel from Min value of a block and store in separate Y bits of every pixel in new mxn block.

The improved decoding steps are organized as follows –

Step1: Read first 128 bits, find which are overhead blocks.

Step2: Except the overhead blocks, take 11 bits block-header and follow the decoding steps described in section 3.2.

The comparative performance in the next section of this paper statistically proves that using 128 bits at the beginning of each block-matrix reduces total image information more than the focused algorithm does. However, these 128 bits can be run-length encoded if necessary.

6 COMPARATIVE PERFORMANCE ANALYSIS

TABLE 1
RESULT OF OVERHEAD BLOCK CALCULATION

Test Image	Total 4x4 Blocks	Overhead Blocks	Total Overhead Bits
Baboon	16384	1926	21186
Lena	16384	3319	36509
Cameraman	16384	6598	72578
Iris	16384	2360	25060

A study over 12,82,048 blocks of 4x4 dimension shows that 73,819 blocks cause overheads. In other words, there are 5.76% blocks for which using 11 bits header of the focused algorithm is meaningless. These 11 bits are redundant for each overhead-block resulting in n x 11 bits that are non-image-information, where n is the number of overhead-blocks. Table 1 shows a portion of our study for some famous test images.

TABLE 2
RESULT OF OVERHEAD BIT REDUCTION BY PROPOSED
MODIFICATION

Test Image	Overhead Bits by Focused Algorithm	Overhead Bits by Proposed Modification	Overhead Bits Reduced
Baboon	21186	16384	4802
Lena	36509	16384	20125
Cameraman	72578	16384	56194
Iris	25060	16384	8676

Our implementation does not preserve 11 bits header for the overhead-blocks as done in focused algorithm. Instead, after dividing a 512×512 image into 128×128 block-matrix, it uses 128 bit long overhead-block-information at the beginning of each of 128 rows. Thus, our improved technique has to use 128×128 bits or 16384 bits instead of 21186, 36509, 72578 and 25060 as overhead bits for the images Baboon, Lena, Cameraman and Iris respectively. A portion of our obtained results is given in Table 2.

As the proposed modification can reduce the overhead bits by a notable extent, the total number of bits after compression by the proposed algorithm is less than that of obtained by focused algorithm. Table 3 shows another comparative study.

TABLE 3
COMPARATIVE PERFORMANCE ANALYSIS

Test Image	Total Bits	Total Bits After Compression by Focused Algorithm	Total Bits After Compression by Proposed Modification
Baboon	2097152	1797568	1792766
Lena	2097152	1282528	1262403
Cameraman	2097152	1181344	1125150

Therefore, if an 8-bit image of 512×512 dimension results in x bits ($x \leq 2097152$) after being compressed by the focused algorithm, our statistical evidence proves that, still, x bits contain p overhead bits, where p is statistically greater than 16384. Thus, preserving 16384 bits instead of p bits results in more compressed image data. However, as 16384 bits required by our modification is kept in 128 different parts, each part can be run length coded as mentioned before.

7 CONCLUSION

In this paper, we investigated a novel image compression technique proposed by Syed & Mehdi [7] and found redundant bits inherently preserved by their technique. Then we suggested an improvement over their algorithm which results in more compression ratio as discussed in our comparative performance analysis. Moreover, further research can be conducted in order to reduce the total number of compressed bits by applying run-length coding on the extra 128 bit binary information suggested by our improvement.

REFERENCES

- [1] Gregory K. Wallace, "The JPEG Still Picture Compression Standard", *IEEE Transactions on Consumer Electronics*, 1991.
- [2] Ralf Steinmetz and Klara Nahrstedt, "Multimedia: Computing, Communications and Applications", 1st Edition, Pearson Education Inc. ISBN: 81-7808-319-1, 2005.
- [3] Suneetha Agarwal and Seshagiri Gurram, "Image Compression using Simple Arithmetic Operations", *International Conference on Computational Intelligence and Multimedia Applications (ICCIMA)*, India, pp. 58-62, 2007.
- [4] Sunil Kumar Pattanik, K. K. Mahapatra and G. Panda, "A Novel Lossless Image Compression Algorithm using Arithmetic Modulo Operation", *IEEE International Conference on Cybernetics & Intelligence Systems (CIS) and Robotics Automation & Mechatronics (RAM) (CIS-RAM 2006)*, Thailand, pp. 234-238, 2006.
- [5] Komal Ramteke and Sunita Rawat, "Lossless Image Compression LOCO-R Algorithm for 16 bit Image", *2nd National Conference on Information and Communication Technology (NCICT)*, pp. 11-14, 2011.
- [6] Rafael C. Gonzalez and Richard E. Woods, "Digital Image Processing", 2nd Edition, Pearson Prentice Hall. ISBN: 81-7758-168-6, 2005.
- [7] Syed Ali Hassan and Mehdi Hussain, "Spatial Domain Lossless Image Data Compression Method", *International Conference of Information and Communication Technologies*, 2011.
- [8] Tinku Acharya and Ajoy K. Ray, "Digital Image Processing: Principles and Applications", John Wiley & Sons, Inc. ISBN: 10 0-471-71998-6, 2005.
- [9] M. Nelson and J. L. Gailly, "The Data Compression Book", 2nd ed. New York: M & T Books, 1996.
- [10] Al-Wahaib and M. S. KokSheikh Wong, "A Lossless Image Compression Algorithm Using Duplication Run Length Coding", *IEEE Conference on Network Application Protocols and Services*, pp. 245-250, 2010.
- [11] J P Cookson and G R Thoma, "X-Ray Image Compression using Run Length Coding", *Journal of Medical Systems*, Volume: 12, Issue: 4, pp. 201-209, 1988.
- [12] Samir Kumar Bandyopadhyay, Tuhi Utsab Paul and Avishek Raychoudhury, "Image Compression using Approximate Matching and Run Length", *International Journal of Advanced Computer Science and Applications (IJACA)*, Volume: 2, Issue: 6, pp. 117-121, 2011.s

A Comprehensive Study of Version Control System in Open Source Software

Nindya Kotwal, Vineeta Bassi

Abstract— The Focus of this paper is to present review about Open Source Software (OSS), Open Source Software Development (OSSD) and how the version control tools are used in OSSD. This paper describes the use of Version Control System (VCS) to coordinate the evolution of open source software projects which mainly includes to manage the communication among various members of a project and to control the hierarchy of the code being developed. The control of the code includes version control and the organization of the team who develops each project.

Keywords— open source software, open source software development process, subversion, version control, version control system.

1 INTRODUCTION

Open Source Software has now a day become the subject of most research and debate; not only for the promise it holds to solve systems development challenges but also due to increase in the OSS development which inturn benefits various organizations in developing nations. Open Source Software development process combines various applicable features found in traditional software processes with some of the other features in such a way that would help in the development and maintenance of high-quality, faster and cheaper software within the rapidly changing Internet environment. Open source development process is not a process without any limitations but it provides numerous applicable advantages and opportunities to the system development process and also to open source society which produces quality softwares very rapidly in open source environment. The one of most important benefit of the open source software is the reduced cost. In addition to this benefit, open source society offers new development models that can be used, particularly within developing nations and economies, and has an extra benefit of skills development. Version control system is an efficient technique that is used in OSS development where work is done in a distributed environment and many version control tools are available.

2 OPEN SOURCE SOFTWARE

2.1 Open Source Software

Open Source Software is the type of software products which are available to the public, with its source code to study, change, and improves its design. However when open source used for commercial purpose, then an open source license is required [1]. Open Source Software is the type of software products which are produced by the distributed communities consisting of huge number of contributors from all over the world who have a sense of commitment towards there work thus in OSS no body is assigning the work and everyone is taking the work willingly which results in skill development.

OSS is software for which the source code is publicly available, though the specific licensing agreements vary as how the code is being used [2]. According to Roets, Minnaar and et al,

there is one basic feature of the Open Source Softwares that results to its four fundamental freedoms and that is, users must have access to the source code. OSS is software whose licenses that is access to code which give users these essential 'freedoms':-

1. To run the program for any of the purpose and also study the workings of the program,
2. To modify the program to suit specific needs and requirements.
3. To redistribute copies of the program at no charge or for a fee.
4. To improve the program, and release the modified version of the program.

2.2 Open Source Development Process

Open source development is an alternative approach used in showing how the Internet can change the way software is initiated, constructed, deployed, and evolved. Open source development offers useful information about common problems as well as some possible solutions for globally distributed product development. Process modeling gives a great opportunity to analyze, compare, visualize, and transfer for reuse these possible solutions [3].

The basic principle for the OSS Development Process (OSSDP) [3] is that by sharing source code and ideas, developers cooperate under a model of systematic peer-reviews, and take advantages of parallel debugging that leads to innovation and rapid advancement in the development of the software [3]. The OSSD model has the following features:

1. Collaborative approach to problem solving through feedback and peer review given by the web community.
2. Large pool of globally dispersed, highly talented, motivated professionals from all over the world.
3. Increased user involvement as users are viewed as co-developers.

The Open Source Software Development model (Fig. 1.) incorporates aspects of various previous models used in the development. In addition, the proposed model attempts to encapsulate the phases of the traditional SDLC and previous OSSD models.

The introduction of a generic initiation phase replaces code phase. The initiation phase can be applied at both the micro and macro level of a project development. This phase refers to developed code that is used as a prototype for further development on a particular project, that code may be a small code segment or the initial version of an entire project. The phase is undertaken either by a developer who develops a piece of code for an existing project, or the project founder in a new project.

The initiation phase then moves into a cycle of code reviews and further contribution from the developer community. The possible number of iterations occurring at this phase is dependent on the interest that project or project related components generates within the developer community. Independent peer review and prompt feedback characterize this phase. The cycle of code contribution and review take place within the wider Internet developer community or web developer community, employing tools such as email, bulletin boards and discussion groups.

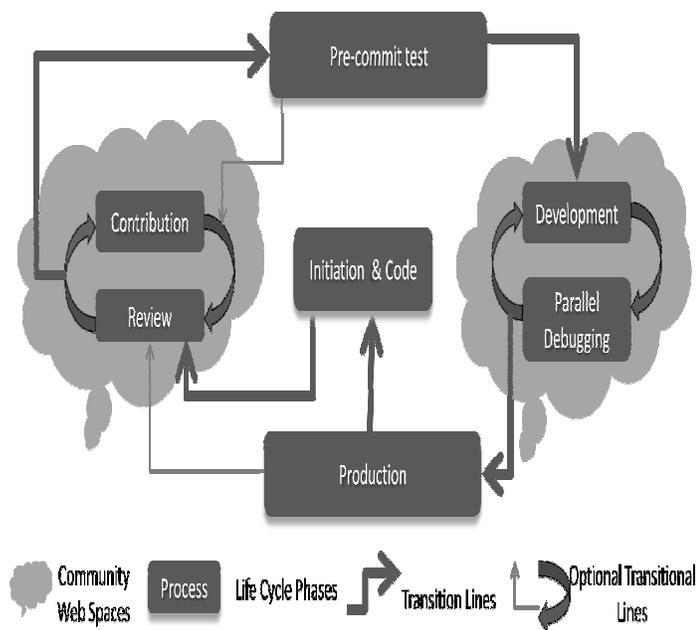


Fig. 1. Open Source Development Model [3]

Once a piece of code is considered adequate for inclusion in a development release, pre-commit testing is performed to ensure that this new piece of code, once added, does not break the existing release. Testing is usually performed by core de-

velopers. No rigorous testing schedule exists and, indeed, the process is not even required. However, the consequences of allowing faulty code into a development release can be severe for the programmers involved and in turn the reputation of the project as a whole.

A process of debugging and reincorporation of code into the development release then takes place. This is again an iterative process occurring within the community web space. No formal planned debugging occurs; individuals volunteer. This is another area exemplifying the strength of open source projects. The more people that seek find and remove bugs, the better the quality of the software which is one of the advantages of the OSSD.

Eventually, code forms part of a production release which is generally managed by core developer. Production releases take the form of a prototype that can be used in the initiation phase of the next iteration of that project, component or code segment.

2.3 Some of the Success Stories of OSS

These are some of the most successful projects developed in open source society.

TABLE 1
Some Examples of Open Source Softwares

GNU/LINUX	Operating System
Apache web server	Web Server
MySQL	Relational Database Management System
OpenOffice	Office Suite
Mozilla Firefox	Web Browser
Thundermail	Email Client
Tcl,Python	Languages
Emacs,Jedit	Text Editors

2.3 Open Source Software Licenses

An OSS license helps publisher in defining the privileges and restrictions that a user must have to follow if they want to use and modify the software. If any one of the developer wants to publish a program as OSS, the publisher can distribute that program as an un-copyrighted product.

TABLE 2
Comparison among OSS licenses [4]

	Notice of modification	Redistribution of modified work	Original Source code attached to modification	Linking to source code	Liability Notice
BSD	Yes	Yes	No	Yes	Yes
GPL	Yes	Only under GPL	Yes	No	Yes
LGPL	Yes	Only under LGPL	Yes	Yes	Yes
MIT	Yes	Yes	Yes	Yes	Yes
APL	Yes	Only under APL	Yes	Yes	Yes

- Nindya Kotwal is currently pursuing masters degree program in Computer science and applications in Thapar University, India.
- Vineeta Bassi is currently working as Assitant Professor in SMCA department in Thapar University, India.

The users of the OSS program can read, copy, modify, and redistribute the program. However, it is also possible for someone to make the program copyrighted by modifying the original program. Consequently, the modified, copyright-protected program becomes a personal property, and is not OSS any more[4].

To prevent this situation, most of OSS licenses implement "copy left" concept: anyone who redistributes the software, with or without changes, must pass along the freedom to further copy and change it [4].

When an OSS program gets published under an Open source software license, the program gets used and distributed by the users in the way as specified in the license. Developers can make their own licenses as well and the procedure of which is given by open source initiative. The most popular set of open source software licenses are those softwares which was approved by the Open Source Initiative (OSI) based on their Open Source Definition (OSD).

3 VERSION CONTROL SYSTEM

3.1 Version Control

A Version Control System (or revision control system) is a combination of technologies and practices for tracking and controlling changes to a project's files, in particular to source code, documentation, and web pages. The reason version control is so universal is that it helps with virtually every aspect of running a project: inter-developer communications, release management, bug management, code stability and experimental development efforts, and attribution and authorization of changes by particular developers [5].

A VCS provides the ability to version resources such as data files. (We use the phrase "a VCS" to mean "an instance of a VCS.") Considerable functionality is associated with modern VCS's; some examples are distributed development, non-linear development and the maintenance of history (versions). A VCS comprises objects such as repositories and branches that help realize such functionality. An authorization scheme for a VCS specifies how these objects are protected, while respecting the semantics of the relationships between the objects.

Version Control System ::= { repository }
repository ::= Branch { Branch }
Branch ::= { Folder } { File } { Tag }
Folder ::= { Folder } { File }

A Version Control System in BNF [6]. The notation ::= stands for "comprises," and {} means 0 or more occurrences

3.2 Version Control in OSS

A version control system maintains an organized set of all the versions of files that are made over time. Version control systems allow people to go back to previous revisions of individual files, and to compare any two revisions to view the

changes between them.

In this way, version control keeps a historically accurate and retrievable log of a file's revisions. More importantly, version control systems help several people (even in geographically disparate locations) work together on a development project over the Internet or private network by merging their changes into the same source repository which is the base of open source software.

There are various version control tools available such as concurrent version system (cvs), subversion (svn), bazaar, Git [7].

3.3 Version Control Basics

There are some basic terms used in VCS:-

1. commit: To make a change to the project; more formally, to store a change in the version control.
2. log message: A bit of commentary attached to each commit, describing the nature and purpose of the commit.
3. update: To ask that others' changes (commits) be incorporated into your local copy of the project; that is, to bring your copy "up-to-date".
4. repository: A database in which changes are stored.
5. checkout: The process of obtaining a copy of the project from a repository
6. working copy: A developer's private directory tree containing the project's source code files, and possibly its web pages or other documents.
7. revision, change, changeset: A "revision" is usually one specific incarnation of a particular file or directory.
8. diff: A textual representation of a change.
9. tag: A label for a particular collection of files at specified revisions.
10. branch: A copy of the project, under version control but isolated, so that changes made to the branch don't affect the rest of the project, and vice versa, except when changes are deliberately "merged" from one side to the other.
11. merge: To move a change from one branch to another.
12. conflict: All version control systems automatically detect conflicts, and notify at least one of the humans involved that their changes conflict with someone else's.
13. lock: A way to declare an exclusive intent to change a particular file or directory.

Version control system is used to organize software projects. A real life example of managing windows to show the working of the version control system is shown in Fig. 2. [8].

There's a main line with stable builds of Windows. Each group (Networking, User Interface, Media Player, etc.) has its own branch to develop new features. These are

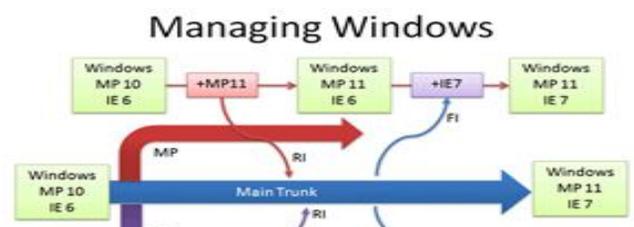


Fig. 2.Managing Windows using VCS

Media Player team makes version 11 in their own branch. When it is ready and tested, there is a patch from 10 – 11 which is applied to Main. This is reverse integration, from the branch to the trunk. The IE team can do the same thing. Later, the Media Player team can pick up the latest code from other teams, like IE. In this case, Media Player forward integrates and gets the latest patches from main into their branch. So it is RI and FI. This arrangement lets changes percolate throughout the branches, while keeping new code out of the main line. In reality, there are many layers of branches and sub-branches, along with quality metrics that determine when you get to RI. But you get the idea: branches help manage complexity. Now we know the basics of how one of the largest software projects is organized using version control system. Subversion is a VCS tool that is used in OSSD.

3.4 Subversion

Subversion is a free/open source version control system (VCS). That is, Subversion manages files and directories, and the changes made to them, over time. This allows you to recover older versions of your data or examine the history of how your data changed. In this regard, many people think of a version control system as a sort of “time machine.” [9].

Subversion can operate across networks, which allows it to be used by people on different computers. At some level, the ability for various people to modify and manage the same set of data from their respective locations fosters collaboration. Progress can occur more quickly without a single conduit through which all modifications must occur. And because the work is versioned, you need not fear that quality is the trade-off for losing that conduit—if some incorrect change is made to the data, just undo that change. When working with the data on a daily basis, developers won't be able to copy, move, rename, or delete files the way they usually do. Instead, they will have to do all of those things through Subversion. “Subversion is an open source version control system” that by default uses a command line user interface on the client side.

4 CONCLUSION AND FUTURE WORK

Open Source Software is software whose source code may be freely studied, modified and redistributed with few restrictions. Version Control is used in Open Source Software as a tool to manage distributed developers all over the world using the common files.

The data are stored in a Version Control System as configuration items, which are usually files. The most common configuration item is a source code file in Open Source Software

Projects. When configuration items are modified, a new revision of the configuration item is created. Configuration items can have multiple modifications between versions, which are published revisions. However, one version of the configuration item can be incompatible with some target systems, so several parallel revisions and versions can be kept in branches. The branches are parallel development lines of the configuration items and software. The source code of the configuration item is duplicated, so different branches can be developed or maintained separately.

Our study provided an evaluation of OSS version control process from the viewpoint of developers distributed all over the world. The research papers and journals were used to produce more understanding about version control process and its basics and how the tools are used to find the evolution of the open source softwares.

The material presented in this paper is only the preliminary review study of the open source software and the importance of version control system in the Open Source Development. In the future the aim is to study the evolution of Open Source Software with large versions, over a period of time using the version control procedures & tool by using a version control tool that is subversion.

REFERENCES

- [1] E.E. Kim, “An Introduction to Open Source Communities,” Blue Oxen Associates Technical report, BOA-00007, 2003.
- [2] Margaret Elliott, Mark S. Ackerman, Walt Scacchi, “Knowledge work artifacts: kernel cousins for free/open source software development,” Proceedings of the 2007 international ACM conference on Supporting group work, November, 2007.
- [3] R. Roets, M. Minnaar, K. Wright, “Towards Successful Systems Development Projects in Developing Countries”, Proceedings of the 9th International Conference on Social Implications of Computers in Developing Countries, Sao Paulo, Brazil, May 2007.
- [4] A. Beard, H. Kim, “A survey on open source software licenses”, Journal of Computing Sciences in Colleges, vol. 22 no. 4, April 2007.
- [5] B. C. Sussman, B. W. Fitzpatrick, C. M. Pilato, “Version Control with Subversion For Subversion 1.6”, 2010, ISBN 9781440495878.
- [6] S. Chamarty, H.D.Patel, M. V. Tripunitara, “An Authorization Scheme for Version Control Systems”, SACMAT'11, Innsbruck, Austria, June 15–17, 2011.
- [7] P. Weibgerber, M. Pohl, M. Burch, “Visual Data Mining in Software Archives to Detect How Developers Work Together,” Computer Science Department, University of Trier, Germany, IEEE Fourth International Workshop on Mining Software Repositories 2007.
- [8] Khalid Azad, “A Visual Guide to Version Control,” Available: <http://betterexplained.com/articles/a-visual-guide-to-version-control>.
- [9] Ben Collins-Sussman, Brian W. Fitzpatrick, C. Michael Pilato “Version Control with Subversion For Subversion 1.5” (Compiled from Revision 2983 Version 1.4.4), 2008, ISBN 9780596510336.

Fuzzy Logic solution for Unit Commitment

Mr.Rudresh.B.Magadum, Smt.G.Suchitra, Dr.S.H.Jangamshetti

Abstract--This paper presents fuzzy logic solution to unit commitment and economic dispatch. Unit commitment is aimed to a proper generator commitment schedule for a power system over a period of one day to one week. The main objective of unit commitment is to minimize the total production cost over the study period and to satisfy the constraints imposed on the system such as power generation-load balance, spinning reserve, operating constraints, minimum up time and minimum down time, etc. In this, fuzzy logic approach is described which achieves a logical and feasible economic cost of operation of power system without the need of exact mathematical formulation

Index Terms— Fuzzy logic, unit commitment, defuzzification fuzzy rules.

◆

1 INTRODUCTION

The use of fuzzy logic has received increased attention in recent years because of its usefulness in reducing the need for complex mathematical models in problem solving. Rather, fuzzy logic employs study of terms which deal with the causal relationship between input and output variables. For this reason, the approach makes it easier to manipulate and solve many problems, particularly where the mathematical model is not explicitly known, or is difficult to solve. Furthermore, fuzzy logic is a technique which approximates reasoning, while allowing decisions to be made efficiently.

2 EXISTING METHODS [5]

2.1 Priority list method

A simple shutdown rule or priority list schemes could be obtained after an exhaustive enumeration of all units combinations of each load level. Priority list methods are easy fast but they are highly heuristic and give schedules with relatively highly production cost.

2.2 Lagrangean relaxation method

In this method minimization cost function decomposed into N smaller minimization problems one for each unit takes more time for solution by iteration procedure. And in this method used to concept of equal incremental cost which will also give high production cost.

2.3 Dynamic programming method

The main limitation of the dynamic programming method is in treating time dependent constraints such as unit minimum up and down time, start-up cost etc. It causes the dimensionality and gives sub-optimal solution.

3 UNIT COMMITMENT PROBLEMS

Shifting load demands on the power supply system require that a sufficient number of generating units be committed to supply the required load. Because of the tremendous expense involved in unit commitment, the electric utility must determine which generators are the most economical to operate and the combinations of units that should be committed to meet a given load demand.

Problems associated with unit commitment have generally been difficult to solve because of the uncertainty of particular aspects of the problem [1]. For instance, the availability of fuels, imprecise load forecasts variable costs affected by the loading of generating units of different fuels or water rates, and losses caused by reactive flows are some of the unpredictable issues. These and other problems of inconsistency affect the overall economic operation of the electric power system. In order to reach a feasible solution to this economic enigma, different constraints must be considered, such as spinning reserve, thermal unit constraints, must-run units, fuel constraints, power generation-load balance, and other operating constraints.

Must-run constraints [5] are included because those units placed on must-run status during particular periods throughout the year are needed as voltage support on the transmission network or to supply process steam. Further, fuel constraints are essential in deriving an optimal energy mix in order to achieve the lowest total fuel cost.

Fuzzy logic [4] represents an effective alternative to conventional solution methods it attempts to quantify linguistic terms so that the variables can be treated as continuous rather than as discrete. In this process, the qualitative behavior of a system, the system's characteristics and response may be described without the need for exact mathematical formulations

4 FUZZY-LOGIC-BASED UNIT COMMITMENT

FUZZY logic is a mathematical theory, which encompasses the idea of vagueness when defining a concept mining [2]. For example, there is uncertainty are 'fuzziness' in expressions like 'large' or 'small', since these expressions are imprecise or relative variables considered thus are termed 'fuzzy.' Fuzziness is simply one means of describing uncertainty. Such ideas are readily applicable to unit problem.

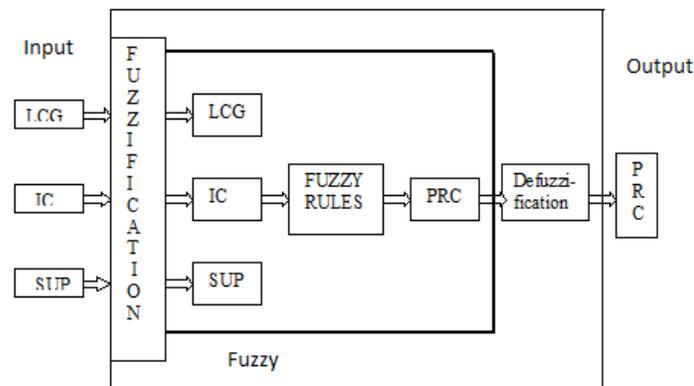


Fig.1 The overall block diagram of unit commitment using fuzzy logic

LCG: Load Capacity of Generator;

IC: Incremental Cost

PRC: Production Cost.

SUP: Start Up Cost;

4.1 Fuzzy Variables

Fuzzy input variables

- Load capacity of generator(LCG)
- Incremental cost(IC)
- Start up cost(SUP)

Fuzzy output variables

- Production cost (PRC)

Load capacity of the generator is considered to be fuzzy, as it is based upon load to be served. Incremental cost is taken to be fuzzy, because the cost of fuel for each unit may be different. Further, the startup cost of unit is assumed to fuzzy, because some units more time than others to be placed on line. Finally, production cost of the system (which includes no load cost of the system) is treated as fuzzy variable since it is directly proportional to the hourly load. Certain other variables, such as minimum up time, minimum down time and generator limitation, are considered to be crisp variables in the unit commitment problem.

4.2 Fuzzy Sets Associated with Unit Commitment [4]

After identifying the fuzzy variables associated with the unit commitment, the fuzzy sets defining these variables are selected and normalized between 0 and 1. This normalized value can be multiplied by a selected scale factor to accommodate any desired variable. The sets defining the load capacity of generator (LCG) are as follows

LCG (MW) = {Low (Lo), Below Average (Bav), Average (Av), Above Average (AAV), High (H)}

The incremental cost (IC) is stated by the following sets,

IC (Rs) = {Least, Small, Large}

The Startup cost (SUP) is defined by the following sets,

SUP (Rs) = {Low, Medium, High}

The production cost, chosen as the objective function is given by,

PRC (Rs) = {Low (Lo), Below Average (Bav), Average (Av), Above Average (AAV), High (H)}

Suitable ranges are selected for the fuzzy sets selected from the given problem

4.3 Membership Function

Based on the fuzzy sets, the membership functions are chosen for each fuzzy input and output variables. Triangular membership function is chosen for all the fuzzy variables.

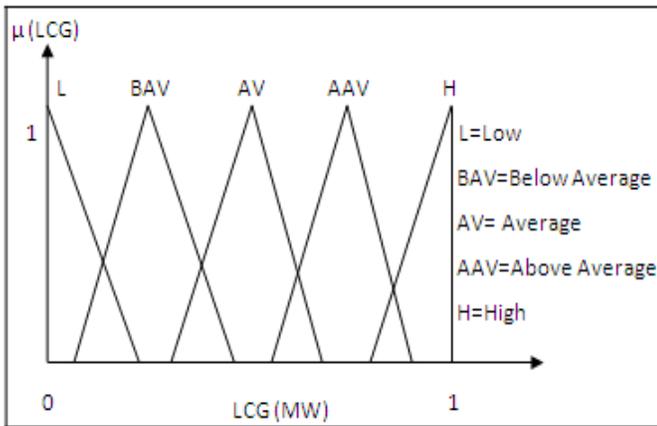


Fig.2 Membership function of load capacity of generators

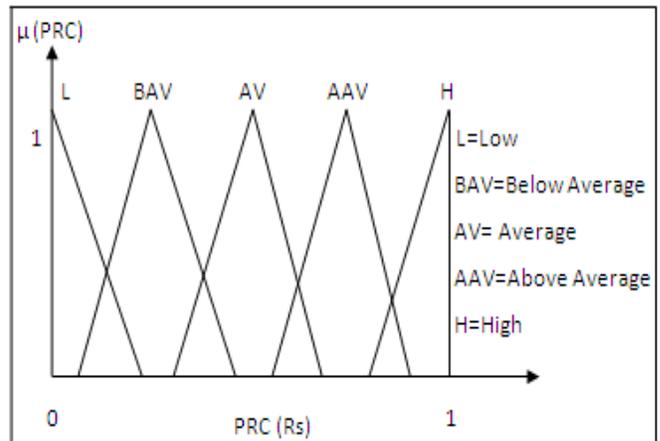


Fig.5 Membership function of production cost

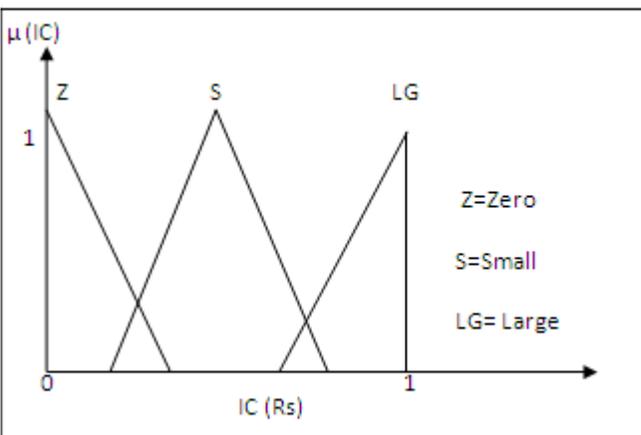


Fig.3 Membership function of Incremental Cost

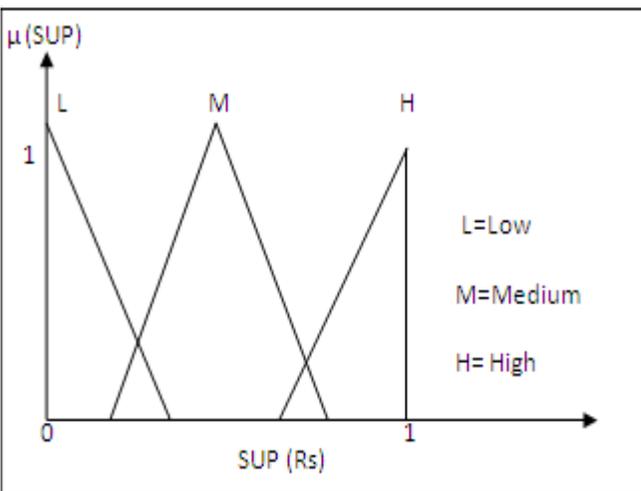


Fig.4 Membership function of Startup cost

Fuzzy If - Then Rules

In a fuzzy-logic-based approach, decisions are made by forming a series of rules that relate the input variables to the output variable using if-then statements. The If (condition) is an antecedent to the Then (consequence) of each rule. Each rule in general can be represented in the following manner:

If (antecedent) Then (consequence)

Load capacity of generator, incremental cost, and startup cost are considered as input variables and production cost is treated as the output variable. This relation between the input variables and the output variable is given as:

Production Cost = {Load Capacity of Generator) and {Incremental Cost} and (start-up Cost)

In fuzzy set notation this is written as

$$PRC = LCG \cap IC \cap SUP$$

Using the above notation, fuzzy rules are written to associate fuzzy input variables with the fuzzy output variable. Based upon these relationships, and with reference to Figs. 2-5, a total of 45 rules can be composed (since there are 5 subsets for load capacity of generator, 3 subsets for incremental cost, and 3 subsets for start-up cost (5*3*3 = 45)). Fig. 6 shows the relationships for some of the rules, and can be applied to all 45.

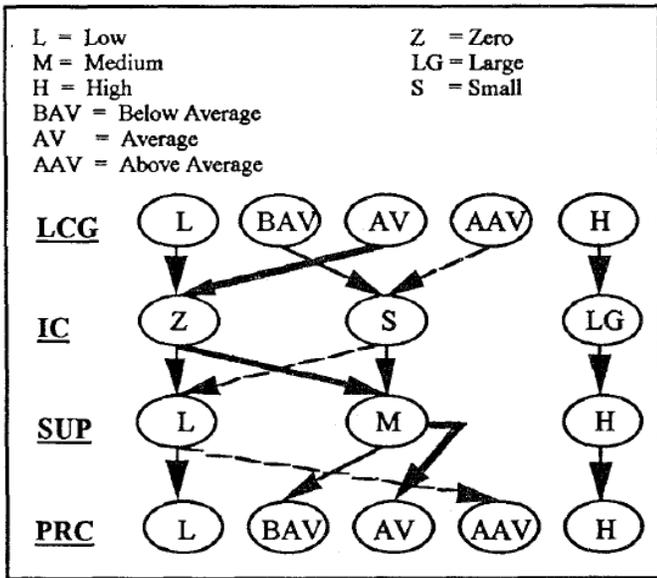


Fig. 6. Relating Input Variables to Output.

These rules are composed in the following manner.

If LCG is (-), IC (-) and SUP cost is (-), then production cost is (-). For example, using figure above 5 rules for the LCG can be written as follows;

Rule 1

If Load Capacity of Generator is Low,

Incremental Cost is Least and

Start up cost is Low, then the production cost is Low.

Rule 2

If Load Capacity of Generator is BAV,

Incremental Cost is small and

Start up cost is Medium, and then the production cost is BAV.

Rule 3

If Load Capacity of Generator is AV,

Incremental Cost is Least and

Start up cost is Medium, and then the production cost is AV.

Rule 4

If Load Capacity of Generator is AAV,

Incremental Cost is small and

Start up cost is Low, and then the production cost is AAV.

Rule 5

If Load Capacity of Generator is High,

Incremental Cost is Large and

Start up cost is high, and then the production cost is high.

In similar manner total 45 rules can be formed.

5 CASE STUDY AND SIMULATION RESULTS

Fuzzy logic simulations are obtained through MATLAB. The result obtained by the fuzzy logic approach (FLA) is compared with solution obtained from Lagrangean Relaxation Method.

TABLE 1
THREE UNIT SYSTEMS

S.N	Pmax (MW)	Pmin (MW)	a (Rs/hr)	b (Rs/MW hr)	c (RS/MW ² hr)
1	450	200	500	5.3	0.004
2	350	150	400	5.5	0.006
3	225	100	200	5.8	0.009

TABLE 2
LOAD IN MW

SL.NO	Load(MW)
1	848
2	890
3	931
4	966
5	1025
6	1000
7	950

TABLE 3
COMPARISON BETWEEN PRODUCTION COST IN RS OBTAINED FROM FUZZY LOGIC WITH LR METHOD

SL.NO	LOAD (MW)	Unit combination	Production Cost in Rs	
			LR Method	Fuzzy Method
1	848	111	7095	7080
2	890	111	7463	7520
3	931	111	7830	7570
4	950	111	8004	7970
5	966	111	8147	8220
6	1000	111	8474	8220
7	1025	111	8716	8220
TOTAL			Rs 55,730	Rs 54,800

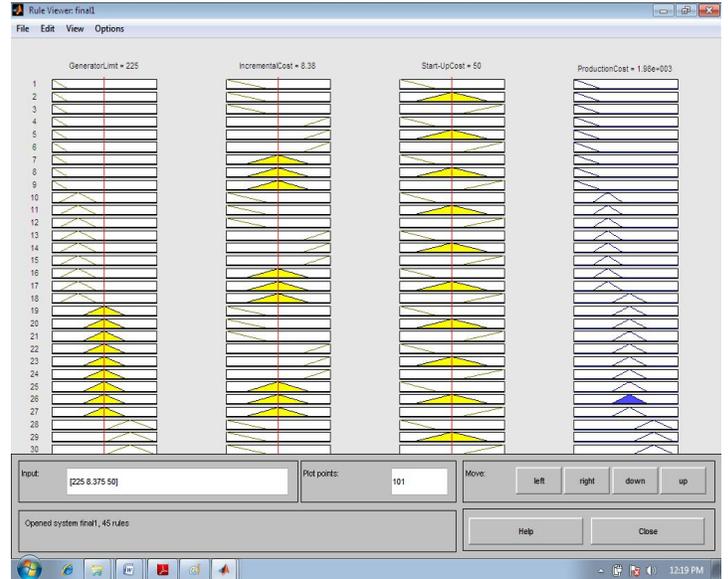


Fig.8 Output of fuzzy in terms of production cost

5.1 Simulation Results

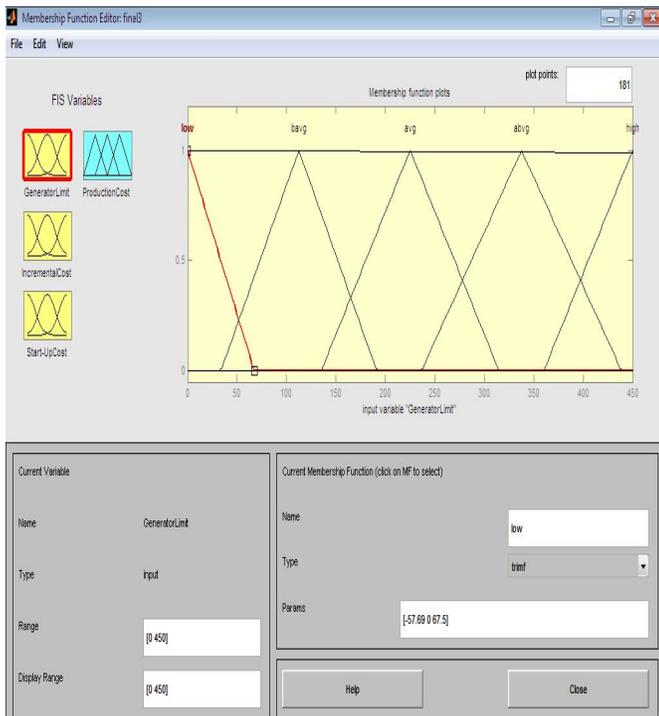


Fig.7 Feeding values to the variables

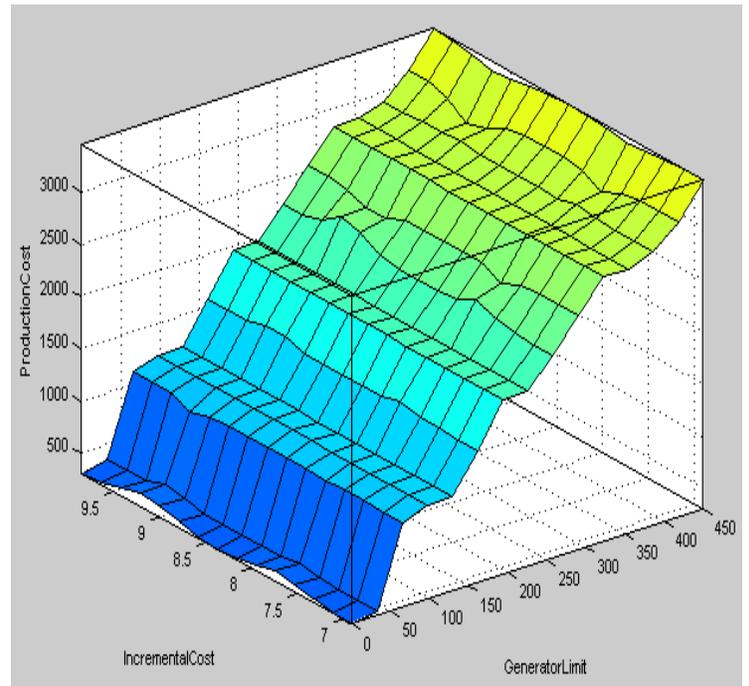


Fig.9 Incremental Cost Vs Production cost considering Generator limit

6 CONCLUSION

The unit commitment and economic dispatch calculations can be solved by using fuzzy logic and this method can be applied to any number of units, each with different operating costs from this approach; we can also conclude that the outcomes are easily understood in terms of the logical representation of the rules. The feasibility of using this method to solve a combinational problem has been demonstrated. The unit commitment schedule is obtained by considering equal incremental criterion. The unit commitment can be described linguistically then such linguistic description can be translated to a solution that yields similar results compared to conventional method.

REFERENCES

- [1] Ahmed Yusuf Saber, Tomonobu Senjyu, Naomitsu Urasaki "Fuzzy Unit Commitment Scheduling Using Absolutely Stochastic Simulated Annealing" IEEE transactions on power systems, VOL. 21, NO. 2, MAY 2006, pp 953-964.
- [2] Dr. K. Duraiswamy, S. Chenthur Pandian "Fuzzy Logic Implementation for Solving the Unit Commitment Problem" 2004 International Conference on Power System Technology - POWERCON 2004 Singapore, 21-24 November 2004, pp 413-418
- [3] D.P.Kadam, P. M. Sonwane, V. P. Dhote, B. E. Kushare "Fuzzy Logic Algorithm for Unit Commitment Problem" International Conference on "Control, Automation, Communication and Energy Conservation -2009, 4th-6th June 2009.
- [4] Seyedrasoul, Saneifard Nadipuram, R. Prasad Howard and Smolleck "A Fuzzy Logic approach to Unit Commitment" IEEE Transactions on Power Systems, Vol. 12, No. 2, May 1997, pp 986-995.
- [5] Allen J Wood and Bruce F Wallenberg, "power generation, operation and Control", John Wiley and sons, New York.

BIOGRAPHY



Rudresh.B.Magadum was born in Karnataka, India on 09 July 1987. He obtained B.E (Electrical and Electronics) from Vishveswaraya Technological University, Belgaum, and Karnataka, India. He is currently pursuing M.Tech Degree in Power and Energy Systems in Electrical and Electronics Engineering, Basaveshwar Engineering College, Bagalkot, India.



Smt.G.Suchitra was born in Bagalkot, Karnataka, India. On 10th May 1960. She obtained B.E degree in Electrical & Electronics from Basaveshwar Engineering College, Bagalkot. She did her post graduation at U.V.C.E. Bangalore University, Bangalore in the year 1990. Presently she is serving as an Assistant professor in the Department of Electrical and Electronics at Basaveshwar Engineering.



Dr. Suresh. H. Jangamshetti: (S'88, M'90, SM'97) was born in Bijapur, Karnataka, India on May 28, 1963. He obtained his B.E (Electrical) degree from Karnataka University Dharwad in 1985 and M.Tech. (Power Systems) & PhD (Wind Energy Systems) from IIT Kharagpur in 1989 & 2000 respectively.

His areas of interest include Wind-Solar Energy Systems, Energy Conservation, Computer Applications to Power System and FACTS He won the "Outstanding IEEE Student Branch Counselor" award for the year 1996(R10) and 2010 (IEEE Bangalore Section) at Basaveshwar Engineering College, Bagalkot, Karnataka, India. He was Fulbright-Nehru Visiting Lecture Fellow at Michigan Technological University, Houghton MI USA during fall 2011. He is working as Professor in the department of E&E at Basaveshwar Engineering College, Bagalkot.

Design and development of a C-Band frequency up-converter system

Anupam Bhardwaj, G.Boopalan, Pinaki Sen, Nupur Sood

Abstract: In this paper a broadband up-converter system has been designed and realized using ADS simulation and designing tool. It is a Radio Frequency transmitter block up-converting the Input Frequency in two stages to reach to the desired frequency level and then transmit it to the antenna port. It contains of oscillators, amplifiers, step attenuators, isolators, filters, and mixer blocks. The input frequency (IF) band is (100 ± 10) MHz and the output frequency band (RF) is (5500 ± 100) MHz. The attenuators are used to increase the dynamic range of the circuit. It is a two stage up-conversion circuit i.e. mixing the IF band at two stages to give the specified RF range. Here two phase locked oscillators are used at the LO port of the mixers, one giving a single frequency and the other one is a programmable frequency synthesizer providing 10 equally spaced frequency channels within S-Band. The frequency synthesizers are designed and simulated using ADS tool and loop filter components are realized using the MATLAB tool.

Key Words: Broadband, phase locked loop, frequency synthesizers, voltage controlled oscillators, mixer conversion loss, frequency divider.

1. INTRODUCTION

In this paper, a C-Band frequency up-converter has been designed and simulated with help of Advanced Design Software (ADS) tool. It is a Radio Frequency transmitter block up-converting the Input Frequency in two stages to reach to the desired frequency level and then transmit it to the antenna port. It contains of oscillators, amplifiers, step attenuators, isolators, filters, and mixer blocks.

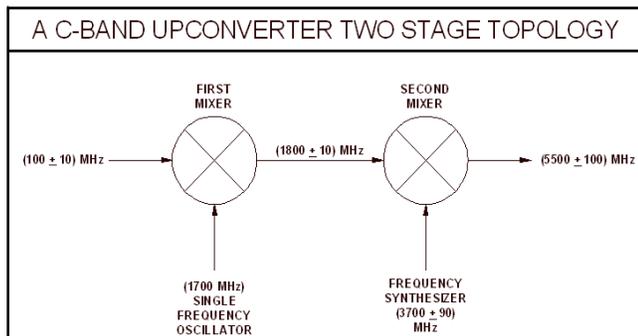


Figure 1 : Frequency Changes in Two Stages Up-converter Block

The input frequency (IF) band is 100 MHz and the output frequency band at radio frequency (RF) is 5500 MHz. The input signal is amplified and converted from 100 MHz to 5500 MHz. The attenuators and amplifiers are used to increase the dynamic range and gain adjustments of the circuit. In this paper, the up-converter selects either of the

dual Payload IF signal input in the frequency band of 100 MHz using a SPDT switch and then up-converts it to the 1st IF using a double-balanced mixer driven by a 1700 MHz Phase Locked Oscillator. The 1st IF is then band-limited to 1800 MHz using a band-pass filter, before being applied to the second mixer, whose Local oscillator is realized through a programmable frequency synthesizer at 3700 MHz, tunable in fixed step size so as to provide 10 channels. The final signal in the band of 5500 MHz is then filtered and suitably amplified before being made available on the front panel of the RF Transceiver. This Frequency synthesizer at the second stage mixing is a phase locked oscillator circuit giving 10 frequency channels with fixed channel spacing at 3700 MHz.

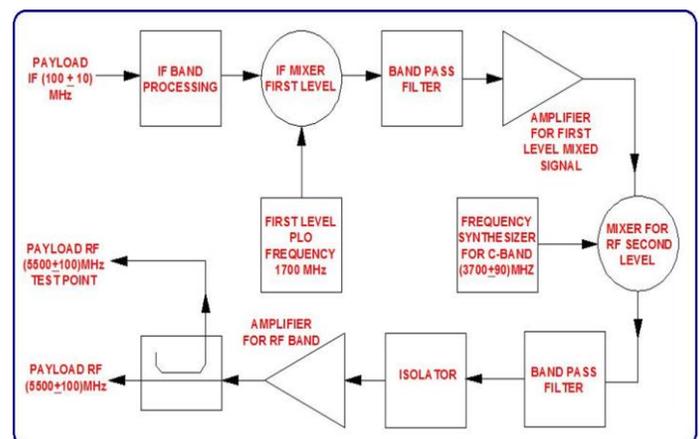


Figure 2 : C-Band Up-Converter Block diagram

This frequency synthesizer circuit consists of a reference oscillator, a reference frequency divider ($R=1$), a phase/frequency detector with a charge pump, a passive low pass higher order filter, a voltage controlled oscillator and a linear VCO frequency divider.

- Anupam Bhardwaj is currently pursuing masters degree program in Communication engineering in V.I.T University, Vellore, Tamil Nadu, India, Mob: +919027884587. E-mail: anu.bhardwaj1@gmail.com
- G.Boopalan is currently working as Sr. Assistant Professor in VIT University, Vellore, Tamil Nadu, India. E-mail: boopalan@vit.ac.in
- Pinaki Sen is currently working in DEAL-DRDO, Dehradun as Scientist-F in the Microwave Division.
- Nupur Sood is currently working in DEAL-DRDO, Dehradun as Scientist-B in the Microwave Division.

In this paper, basically two design iterations are employed with the difference in the second iteration is the presence of the programmable frequency synthesizer rather than a simple oscillator giving 10 channels as used in the first iteration. Also in the second iteration minor changes in the parameters of the various components are done to get a better output level and better level of up-conversion. And hence both the design iterations are compared on the basis of output power level, level of mixing of the signal and difference of level at required frequency and spur offset frequency.

2. SPECIFICATIONS

The block diagram and target specification for the up-converter is shown below. Two design iterations were performed. During the second iteration, minor changes are done and a frequency synthesizer circuit is added in the place of simple 10 channel generating oscillator at the input to second mixer as in the first iteration.

PARAMETERS		SPECIFICATIONS REQUIRED
Input	- Frequency Band - Power Level	(100 ± 10) MHz (0 ± 2) dBm nominal
Output	- Frequency Band - Power Level	(5500 ± 100) MHz (0 ± 2) dBm nominal
Output RF channel selection		Through frequency synthesizer programming via M&C port
Spurious Response		≤ -100 dBc/Hz within the output RF band
Frequency Synthesizer for Second Mixer	- Output - Number of channels - Channel Spacing - Phase Noise	(3700 ± 90) MHz 10 (Ch:1-3610 MHz; Ch:10-3790 MHz) 20 MHz ≤ -75 dBc/Hz at 1 kHz offset ≤ -100 dBc/Hz at 10 kHz offset ≤ -120 dBc/Hz at 100 kHz offset

3. VARIOUS SUBCIRCUITS

AMPLIFIERS:-

There are three amplifiers are used in the up-converter circuitry. Two of these amplifiers are used in the up-converter chain itself while one is used in the frequency synthesizer output. These amplifiers, Amp1 and Amp3, used in the up-converter are providing gain of 19dB and 15dB respectively.

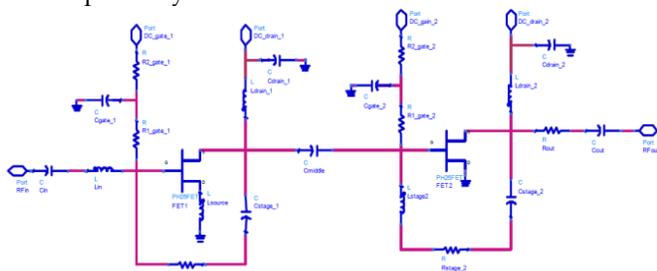


Figure 3 : Amplifier Topology used

While the one, Amp2, used at the frequency synthesizer

output is giving gain of 12dB. Here, the schematic used for getting the required gain is shown. To improve input matching and stability, an inductor was added on the source of the first stage. The inter-stage matching network is simply a capacitor for DC isolation. Using an inductor in the feedback loop for the second stage increased the gain at the higher end of the frequency band. To improve input matching and stability, an inductor was added on the source of the first stage. The inter-stage matching network is simply a capacitor for DC isolation. Using an inductor in the feedback loop for the second stage increased the gain at the higher end of the frequency band.

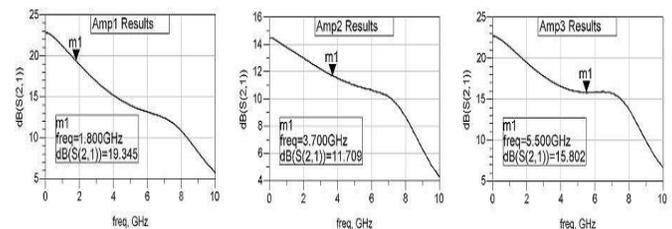


Figure 4 : Respective results for the Amplifiers

ATTENUATORS:-

The attenuator block was realized by a double pi-attenuator. It consists of lumped elements resistors, inductors and capacitors. Here inductors and capacitors are used at the input and output ports for coupling while the resistors R1 and R2 are responsible to give attenuation values.

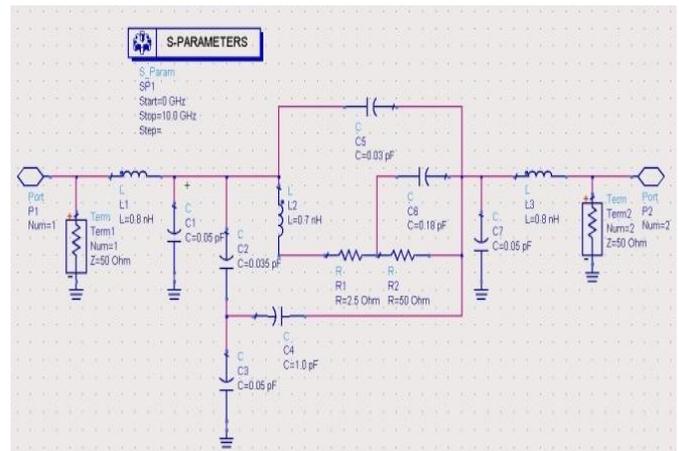


Figure 5 : Attenuator circuitry in ADS

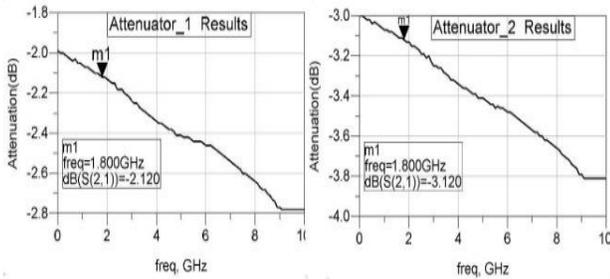


Figure 6 : Attenuation results in ADS

The attenuation of 2dB and 3dB are required here to increase the dynamic range of the up-converter. These attenuators circuits used here are Pi-pad attenuators and the attenuation is depending upon the resistors values.

MIXERS:-

Here two single balanced diode mixers are used in the two stage up-converter topology. The first mixer is used to up-convert the input of (100 ± 10) MHz to (1800 ± 10) MHz using 1700 MHz at its LO input. It provides a conversion loss of 6 dB and it needs a LO power level to be 10dBm. The second mixer is used for up-converting the input of (1800 ± 10) MHz to the required RF output (5000 ± 100) MHz using the frequency synthesizer giving (3700 ± 90) MHz at the LO input of the second mixer. It provides a conversion loss of 6 dB and it needs a LO power level to be 10dBm.

FREQUENCY SYNTHESIZER:-

There is a frequency synthesizer employed at the LO input of the second mixer. The frequency synthesizer has the following specifications. This frequency synthesizer is based on the concept of phase locked oscillator (PLO). Phase – locked – loop (PLL) is a feedback system used to lock the output frequency and phase to the frequency and phase of input. It can also be used as a frequency synthesizer for modulation and demodulation. Normally, PLL is used in both transmitting and receiving terminals for any wireless communication. It is composed of several components, which are phase detector, filter, voltage controlled oscillator and main divider. It operates as a negative feedback loop. While the voltage controlled oscillator generates an output signal, its output phase is fed and compared to the reference signal. This process continues until no phase difference exists. At this state, PLL is called “frequency locked”. Noise in PLL is classified into two categories, which are amplitude noise and phase noise. Amplitude noise is detected and terminated easily. In contrast, phase noise is difficult to identify and express in an equation due to unpredictable characteristics of electronic components. It affects the system performance and the signal to noise ratio.

Moreover, the benefit of having a more reliable phase noise model deals mainly with the design problem. Here, in the PLO block diagram firstly there is a crystal oscillator giving a reference frequency (**Fref**) followed by a reference frequency divider(**R**). This is one of the inputs to the phase frequency detector having another input as the VCO output (**Fout**) divided by the VCO divider (**N**). The basic need for the phase frequency detector to work is that both the inputs should be at same frequencies

$$F_{ref} / R = F_{out} / N$$

This phase frequency detector works as charge pumps i.e. phase detector (or comparator) is used to convert the phase error (i.e. the phase difference between the two inputs of the phase frequency detector) into the equivalent current (**Icp**) and **K_o** is referred to as charge pump gain constant.

$$I_{cp} = (K_{\phi} / 2\pi) Amp$$

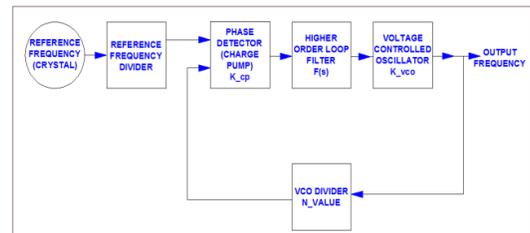


Figure 7 : PLO Block Diagram

PARAMETER	SPECIFICATION REQUIRED
Output Frequency Range	(3700 ± 90) MHz
Step Size	20 MHz (10 channels, Ch 1: 3610 MHz – Ch 10 : 3790 MHz)
Output Power level	+7 dBm
Input Reference	10 MHz sine-wave
- Frequency	-1 to -10 dBm
- Level	± 0.01 ppm
- Stability	
Phase Noise	≤ -75 dBc/Hz at 1 kHz offset ≤ -80 dBc/Hz at 10 kHz offset ≤ -100 dBc/Hz at 100 kHz offset
Power Supply	+3.3V DC, 156 mA

The problem that comes with the designing of the fourth-order loop filter is maintaining the required loop natural frequency (ω_n) and phase margin (PM) when adding the extra attenuation.

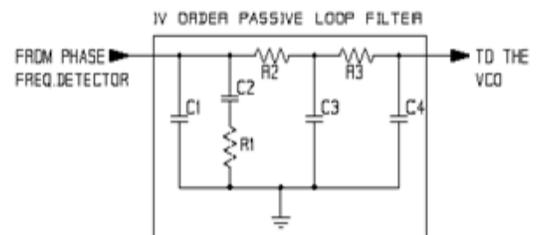


Figure 8 : Loop Filter used

Within this Section a new set of formulas are presented for the exact calculation of all the fourth-order time constants, which guarantee the specified loop phase and gain margins in a PLL. This is achieved by focusing on the behavior of the transfer functions only at the specified loop natural frequency. The loop filter used in the frequency synthesizer circuitry is shown below. This loop filter is a passive fourth order low pass filter which is used for locking the frequency to the required state and providing the required level of attenuation for the spurious responses. Increasingly often a fourth-order loop filter design is used to improve the first reference spur attenuation by adding another RC pole.

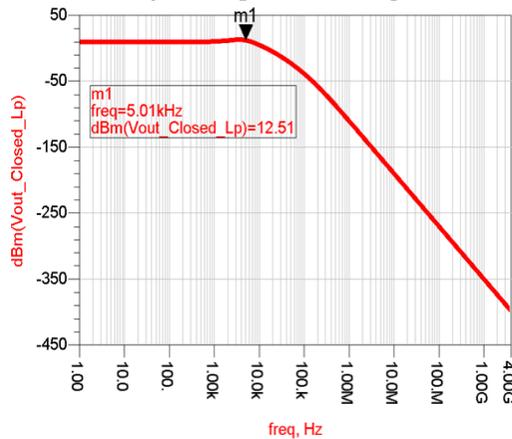


Figure 9 : Closed Loop Frequency response

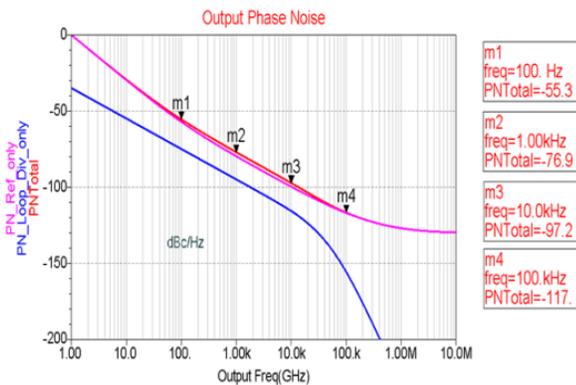


Figure 10 : Phase noise variation for the frequency synthesizer

The specified attenuation frequency could equally be required at another frequency offset, to reduce other known spurious products within the synthesizer architecture. For this analysis, the required attenuation is assumed to be at the first sampling frequency spur offset, from the carrier. The frequency responses for the open and closed loop are obtained for the phase locked loop using MATLAB tool and shown below. The loop filter transfer function is represented as

$$F(s) = \frac{(1+sT2)}{s(1+sT1)(1+sT3)(1+sT4)}$$

The phase margin will be given as:

$$PM = \pi - \tan^{-1}\left(\frac{F(j\omega n)}{N}\right);$$

The time constants are calculated as:

$$T1 = \frac{(\sec PM + \tan PM)}{\omega n}; T2 = \frac{\gamma}{(\omega n)^2 * T1}; T0 = \frac{K\phi * Kv * T1}{N * \omega n}$$

$$T3 = T(3,1) * T1; T4 = T(4,1) * T1$$

T(3,1) and T(4,1) are time constant ratios generally of the order of 10^{-3} or 10^{-4} . Now the loop filter components are calculated using the time constants, natural frequency (ωn), phase margin, charge pump gain constant ($K\phi$), VCO gain constant (Kv) and the VCO divider (N) value.

$$C1 = (T0 * T2) / T1; C2 = T0 - C1; R1 = T1 / C2;$$

$$R2 \cong 2 * R1; R3 \cong 5 * R1; C3 = T3 / R2; C4 = T4 / R3$$

The output phase noise has been calculated and obtained using the ADS tool for the above calculated loop filter components.

4. SIMULATION RESULTS

Here, the two iterations are followed in order to receive the output at the desired radio frequency. In this paper, basically two design iterations are employed with the difference in the second iteration is the presence of the programmable frequency synthesizer at the LO port of the second mixer rather than a simple oscillator giving 10 channels as used in the first iteration. Also in the second iteration minor changes in the parameters of the various components are done to get a better output level and better level of up-conversion. And hence both the design iterations are compared on the basis of output power level, level of mixing of the signal and difference of level at required frequency and spur offset frequency.

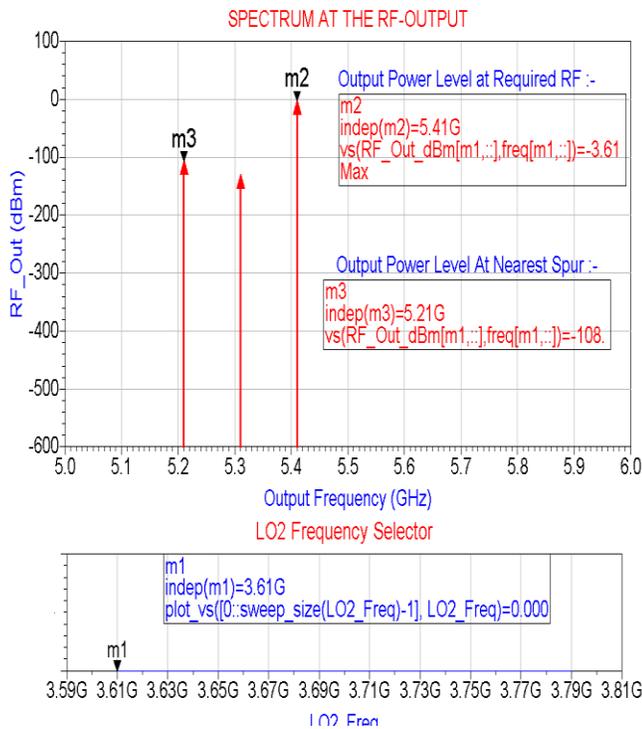


Figure 11 : RF Spectrum for the First iteration

The two iterations has been designed and simulated in ADS and there results are shown below. For the first iteration the local oscillator frequency (LO-2) is taken as swept variable an on changing this frequency according to the channel spacing different output required channels are obtained. And for the second iteration the VCO divider (N) value is changed in order to change the output of the frequency synthesizer and thus referring to the various required output channels. Here figure 11 and 12 shows the output spectrum for the first and second iteration respectively. In figure 13 output spectrum obtained in the second iteration is shown followed by the description of power level of the signal at the various critical stages of the up-converter chain.

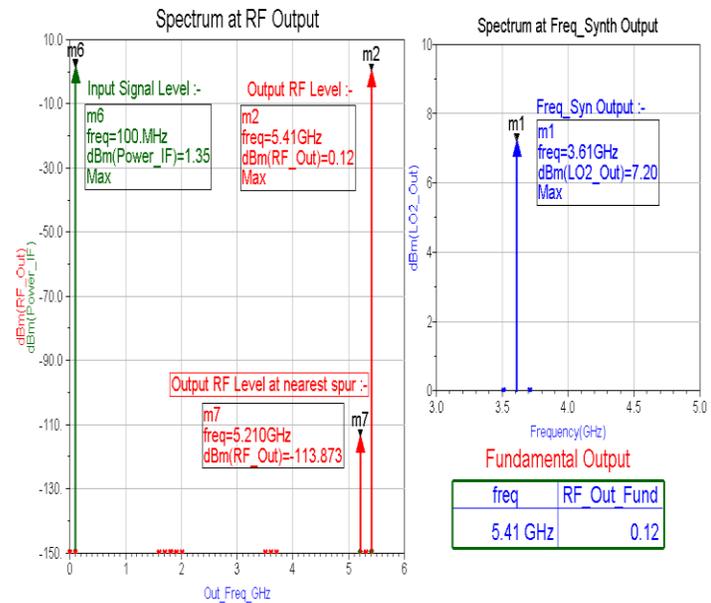


Figure 12 : Output spectrum for the second iteration

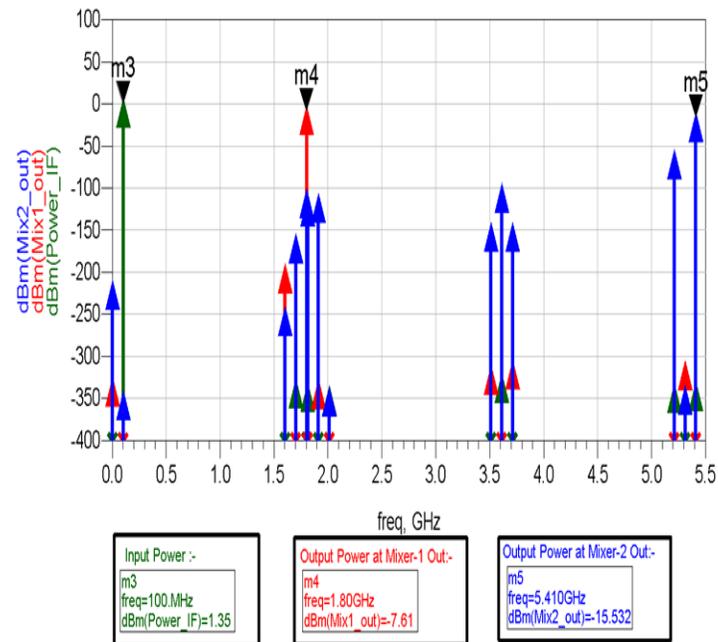


Figure 13 : Power levels at critical stages for the second iteration

There are also shown the measurement result for the frequency synthesizer and the final output for the up-converter portion.

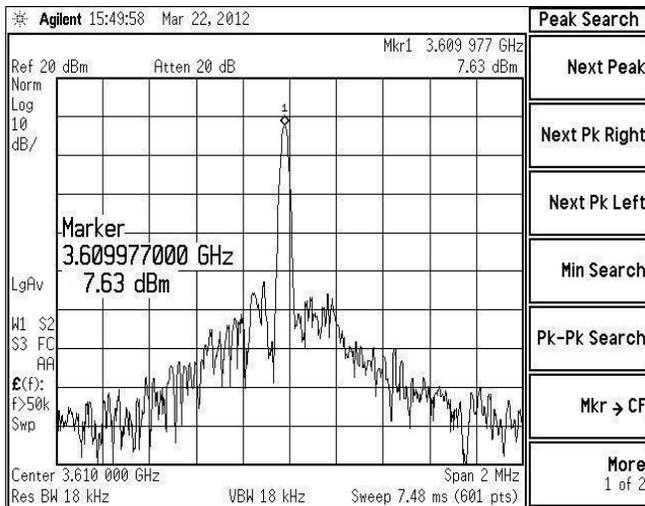


Figure 14: frequency synthesizer output

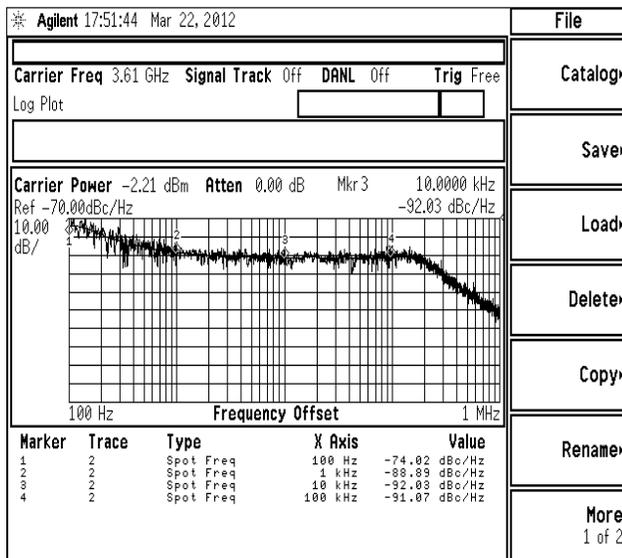


Figure 15: frequency synthesizer output phase noise

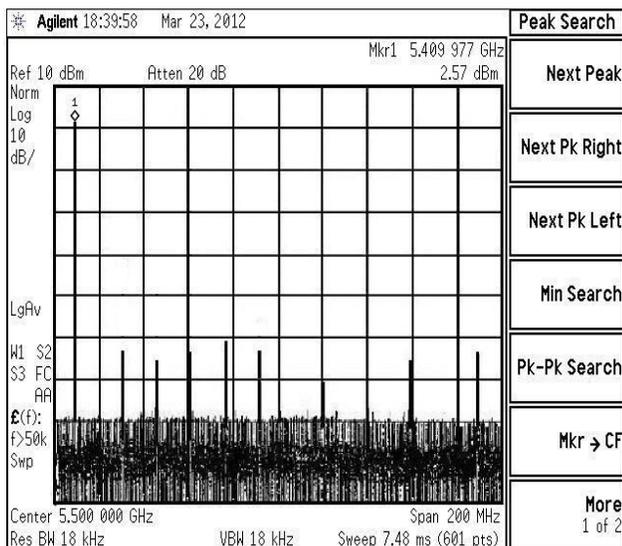


Figure 16: Up-converter measured final output for second iteration

5. CONCLUSIONS

- In figure 13, mixer 1 is giving output at -7.61 dBm with the input at +1.35 dBm thus for the calculation of mixer conversion loss is 8.16 dB.
- In figure 13, mixer 2 is giving output at -15.53 dBm with the input to the mixer 2 is at -7.61 dBm thus for the calculation of mixer conversion loss is 7.9 dB.

Thus, if we take a look at the output spectrum of both the iterations we can see that the output level in the second iteration is improved to the required level and the spurious level is further decreased by 5 to 6 dB. Thus, the signal separation from the spurious level is improved by 10 to 15 dB.

6. REFERENCES

- [1] "An ultra broadband up converter circuit Auden tangs", Stefan Johansson and Johan Lingsten Eriksson microwave systems, Se-431 84 Sweden.
- [2] "C-Band Up and Down Converter Design Report" EEE 787: MMIC design by Ray Gabany, Johns Hopkins University 2000.
- [3] "A Low Distortion C-band Double Balanced Up-converter MMIC" C.H. Lee, S. Chakraborty, and I. Lascar School of ECE, Georgia Institute of Technology Yamacraw Design Center, Atlanta, GA 30332, USA ,2001.
- [4] "A downsized and integrated C-band transceiver for VSAT" Ge Zhiqiang; Microwave and Optoelectronics Conference, 1995. Proceedings, 1995 SBMO/IEEE MTT-S International Volume: 1 Publication Year: 1995 , Page(s): 33 - 36 vol.1
- [5] "The up-converter with digital control for satellite communication stations" Alybin, V.G.; Zyblikov, S.Y.; Parusov, A.A.; E.Y.; Microwave Conference, 1999. Microwave & Telecommunication Technology. 1999 9th International Crimean Digital Object Identifier: 10.1109/CRMICO.1999.815141 Publication Year: 1999, Page(s): 55 - 56 Cited by: 1
- [6] "Simulation of mathematical phase noise model for a phase-locked-loop" 2000 Rangsi Watakapong and Duangrat Eungdamrong Sirindhron International Institute of Technology, Thammasat University.
- [7] " Fourth-order PLL loop filter design technique with invariant natural frequency and phase margin" by Thompson and P.V. Brennan; IEEE Proc.-Circuits Devices Syst., Vol. 152, No. 2, April 2005
- [8] "Design and simulation of Fractional-N PLL frequency synthesizers" By Mücahit Kozak and Eby G. Friedman Department of Electrical and Computer

Engineering University of Rochester

- [9] PLL design by Dean Banerjee, PLL Techniques by Floyd M Gardener, Pozar, D. Microwave Engineering, Second Edition. John Wiley & Sons, 1998.

Power System Economic Dispatch Using Traditional and Neural Networks Programs

Draidi Abdellah, Labeled Djamel.

Abstract— The introduction of techniques of artificial intelligence in software of control and decision is an essential element in research and development of tomorrow's power systems. Neural networks are among the techniques most used in the field of artificial intelligence. The economic dispatch is a key sector in the electricity networks, where it must generate less energy for the same demand with good economic operation reducing repartition grid losses to have the least cost of kWh possible. In this paper, we will opt for a quicker economic dispatch; we will program a mesh network of 8 buses including 3 generation units using traditional program then backpropagation learning neural network program, finally, we will compare the two programs in terms of speed and reliability.

Index Terms— Power systems, economic dispatch, artificial intelligence, neural networks, grid losses, traditional program, backpropagation learning.

1 ECONOMIC DISPATCH

1.1 Introduction

THE Economic dispatch is a static optimization problem which consists in distributing active power production requested by different grid buses from generation unites in the most economical way. This distribution must of course respect the limits of production of generation units. The variable to be optimized is the production cost.

1.2 The cost function

The cost of production of a plant is generally modeled by a polynomial function of second degree in P_{Gi} (active power generated by the plant i) whose coefficients are constants specific to each plant: [1]

$$C_i(P_{Gi}) = a_i + b_i P_{Gi} + c_i P_{Gi}^2 \quad (1)$$

1.3 Economic dispatch solution

To minimise the total production cost of an interconnected power system we must minimize the sum of cost functions of production units

$$\text{Minimise} \quad C = \sum_{i=1}^{ng} C_i(P_{Gi}) \quad (2)$$

taking into consideration the following constraints:

$$\text{Equality constraints:} \quad \sum_{i=1}^{ng} P_{Gi} = \sum_{j=1}^{nd} P_{Dj} \quad (3)$$

$$\text{Inequality constraints:} \quad P_{Gi}^{\min} \leq P_{Gi} \leq P_{Gi}^{\max} \quad (4)$$

where C is total cost function, ng is a total number of producer nodes and nd is a total number of consumer nodes. P_{Gi} represents the active power generated by the i th generator, P_{Dj} is the active power consumed by the j th load, P_{Gi}^{\max} is the maximum active power of the i th generator and P_{Gi}^{\min} is the mini-

mum active power of the i th generator. [2]

The solution of this problem is obtained by using the Lagrange function which is obtained by multiplying the function of equality constraints by the Lagrange multiplier λ , adding to the total cost function (5).

$$L(P_{Gi}, \lambda) = C + \lambda \left(\sum_{i=1}^{ng} P_{Gi} - \sum_{j=1}^{nd} P_{Dj} \right) \quad (5)$$

The derivatives of the Lagrange equation with respect to each independent variable (P_{Gi}, λ) give us: [3]

$$\frac{\delta L}{\delta P_{Gi}} = \frac{dC_i(P_{Gi})}{dP_{Gi}} - \lambda = 0 \Rightarrow \lambda = \frac{dC_i(P_{Gi})}{dP_{Gi}} \quad (6)$$

$$\frac{\delta L}{\delta \lambda} = \sum_{j=1}^{nd} P_{Dj} - \sum_{i=1}^{ng} P_{Gi} = 0 \quad (7)$$

So, from (6) λ represents the incremental cost of the i th generator, then, for each energy packet the generator having the least λ is responsible of production (least cost) respecting the constraints of (3) and (4).

1.4 Insertion of losses formula in the economic dispatch

1.4.1 Calculation of Losses (P_L):

The general formula of losses following the equations of power flows is:

$$P_L = \psi^T G \psi \quad (8)$$

with $\psi = M \delta$. M and δ are matrices of *line's incidence* and *phases of nodes* respectively. G is the diagonal matrix of line conductances.

$$G = \text{diag} [G_{12} G_{13} G_{14} \dots G_{(n-1)n}] \quad (9)$$

δ can be approximated by a DC Load Flow so,

$$P_G - P_D = A \delta \Rightarrow \delta = A^{-1} (P_G - P_D) \quad (10)$$

A represents the DC Load flow Matrix, therefore,

$$P_L = P_D^T B P_D - 2P_D^T B P_G + P_G^T B P_G \quad (11)$$

where

$$B = A^{-1} M^T G M A^{-1}$$

- Draidi Abdellah is currently pursuing Doctorat degree in electric power engineering at the electrical engineering department, and member of Electrical Engineering Laboratory of Constantine, University of Constantine, Algeria, PH-00213551826942. E-mail: abdellah.draidi@gmail.com
- Dr.Labeled Djamel is currently a teacher at the electrical engineering department and member of Electrical Engineering Laboratory of Constantine University of Constantine, Algeria, PH-00213773312865. E-mail: djamel_labeled@yahoo.fr

1.4.2 Penalty Factor: supposing

P_{Gi} Power generated by the i th plant.
 P_{Ci} Part of the power generated that is really consumed by loads.
 P_{Li} Part of the power generated that is lost in the lines, we know that:

$$P_{Gi} = P_{Ci} + P_{Li} \quad (12)$$

and from(1)
$$\frac{dC}{dP_{Gi}} = b_i + 2c_i P_{Gi} \quad (13)$$

by substitutions
$$\frac{dC}{dP_{Gi}} = b'_i + 2c'_i P_{Gi} \quad (14)$$

where $b'_i = b_i f_i$ & $c'_i = c_i f_i$ are the new coefficients of (1), with $f_i = (1 - \frac{dP_{Li}}{dP_{Gi}})^{-1}$ is the penalty factor of the incremental cost.

1.4.3 Criterion of convergence:

If $\left| \sum_{i=1}^{ng} P_{Gi} - \sum_{j=1}^{nd} P_{Dj} - P_L \right| \leq \epsilon$ the system has converged. [1]

2 NEURAL NETWORKS

2.1 Introduction

The origin of artificial neural networks comes from the biological neuron modeling test by Warren McCulluch and Walter Pitts. They assumed that the nerve impulse is the result of a simple calculation made by each neuron and the thought is born with the collective effect of a network of interconnected neurons. [4]

2.2 Neuron Model

A neuron consists essentially of an integrator that performs the weighted sum of its inputs. The result n of this sum is then transformed by a transfer function f which produces the neuron output a . Fig. 1

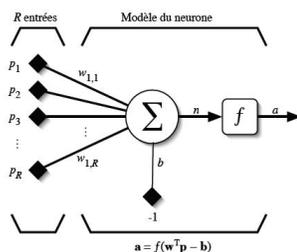


Fig. 1. Artificial neuron model

$$n = \sum_{j=1}^R w_{1,j} p_j - b \quad (15)$$

$$= w_{1,1} p_1 + w_{1,2} p_2 + \dots + w_{1,R} p_R - b$$

This output corresponds to a weighted sum of the weights w_{ij} and inputs p_i minus the bias b . [5]

2.3 Neural network learning

There are essentially two types of learning, unsupervised and supervised. In our paper we will use a supervised learning

where we impose to the network specific operations by forcing it from inputs submitted, the outputs to take by changing the synaptic weights. [6]

2.4 Error backpropagation learning

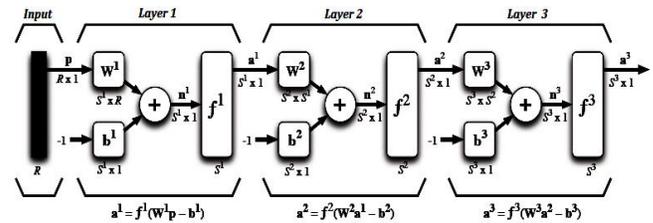


Fig. 2 Representation of three layers network

The simple perceptron consists of a single layer of S neurons which are fully connected to vector p of R entries. In a multilayer perceptron the equation that describes the output of layer k (Fig. 2) is given by:

$$a^k = f^k (W^k a^{k-1} - b^k) \quad \text{for } k = 1, \dots, M \quad (16)$$

where M is the total number of layers. The network outputs correspond to a^M . The backpropagation algorithm uses the mean squared error as performance index, and allows a supervised learning with a set of associations (stimulus, target) $\{(p_q, d_q)\}$, $q=1, \dots, Q$ where p_q represents the stimulus vector (inputs) and d_q the target vector (desired outputs). At each time t , we can forwardpropagate a different stimulus $p(t)$ through the network of Fig. 2 to obtain an output vector $a(t)$. This allows us to calculate the error $e(t)$ between what the network produces as output for the stimulus and the target $d(t)$ associated with it:

$$e(t) = d(t) - a(t) \quad (17)$$

The performance index F minimizes the mean square error. This index is approximated by the instantaneous error $\hat{F}(x)$, the vector x includes all the weights and biases of the network,

$$\hat{F}(x) = e^T(t) e(t) \quad (18)$$

we use the *gradient descent method* to optimize x :

$$\Delta w_{i,j}^k(t) = -\eta \frac{\partial \hat{F}}{\partial w_{i,j}^k}; \quad \Delta b_i^k(t) = -\eta \frac{\partial \hat{F}}{\partial b_i^k} \quad (19)$$

where η represents the learning rate, so:

$$\frac{\partial \hat{F}}{\partial w_{i,j}^k} = \frac{\partial \hat{F}}{\partial n_i^k} \times \frac{\partial n_i^k}{\partial w_{i,j}^k} \quad (20)$$

$$\frac{\partial \hat{F}}{\partial b_i^k} = \frac{\partial \hat{F}}{\partial n_i^k} \times \frac{\partial n_i^k}{\partial b_i^k} \quad (21)$$

n_i^k represent the *activation levels* of a layer k which depend directly on weights and bias on this layer;

$$n_i^k = \sum_{j=1}^{s^{k-1}} w_{i,j}^k a_j^{k-1} - b_i^k \quad (22)$$

so, the second term of (20) and (21) becomes:

$$\frac{\partial n_i^k}{\partial w_{i,j}^k} = a_j^{k-1}; \quad \frac{\partial n_i^k}{\partial b_i^k} = -1 \quad (23)$$

for the first term of (20) and (21), we define the $ty_{s_i^k}$ where $s_i^k = \frac{\partial \hat{F}}{\partial n_i^k}$, then (19) becomes:

$$\Delta w_{i,j}^k(t) = -\eta s_i^k(t) a_j^{k-1}(t); \Delta b_i^k(t) = \eta s_i^k(t) \quad (24)$$

with

$$s^k = \left(\frac{\partial n^{k+1}}{\partial n^k} \right)^T \frac{\partial \hat{F}}{\partial n^{k+1}} = \hat{F}^k(n^k) (W^{k+1})(s^{k+1}) \quad (25)$$

In this case, we get a recursive formula where sensitivity layers upstream (input s^k) depend on the sensitivity of the layers downstream (output s^{k+1}). That's where the term «back-propagation» comes, because the direction of information propagation is reversed compared to that of (16). [5]

3 ECONOMIC DISPATCH USING NEURAL NETWORKS

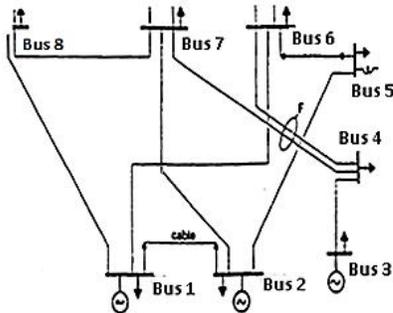


Fig. 3 The 138KV area from IEEE 24-Bus Test Network[7]

In this example we will study and program the economic dispatch of a mesh network using the traditional and the neural network programs. The network is the 138KV area in the IEEE 24-Bus Test Network (Fig. 3) and constituted of 8 busbars; three Generation buses (Bus 1, Bus 2 and Bus 3) and five Load buses (Bus 4, Bus 5, Bus 6, Bus 7 and Bus 8) [7]

First, we show the characteristics of generators;

TABLE 1
CHARACTERISTICS OF GENERATORS [7]

Gen Bus	Pmax [MW]	Pmin [MW]	a [\$ /h]	b [\$ /MWh]	c [\$ /MW²h]
Bus 1	308	0	646.99	19.18	0.0322
Bus 2	350	0	646.99	19.18	0.0322
Bus 3	250	0	1829.71	27.22	0.0628

Second, we give busbars distances and impedances;

TABLE 2
DISTANCES AND IMPEDANCES BETWEEN BUSBARS [7]

Line	from	to	Distance [Miles]	R[Ω]	X [Ω]
1	1	8	55	0.055	0.21
2	1	6	45	0.02	0.08
3	1	2	3	0.003	0.014
4	2	7	60	0.015	0.115
5	2	5	50	0.05	0.192
6	3	4	16	0.016	0.06
7	4	7	43	0.043	0.165

8	4	6	43	0.043	0.165
9	5	6	16	0.014	0.061
10	7	8	31	0.031	0.119

Afterwards, we present energy demands of load buses;

TABLE 3
ENERGY DEMANDS OF LOAD BUSES

DEMAND [MW]					TOTAL DEMAND [MW]
B4	B5	B6	B7	B8	
0	46	1	15	90	152
0	56	10	25	100	191
0	66	20	35	110	231
0	76	30	45	120	271
0	86	40	55	130	311
0	96	50	65	140	351
0	106	60	75	150	391
10	116	70	85	160	441
20	126	80	95	170	491
30	136	90	105	180	541
40	146	100	115	190	591
50	156	110	125	200	641
60	166	120	135	210	691
70	176	130	145	220	741
80	186	140	155	230	791
90	196	150	165	240	841
100	206	160	175	250	891

3.1 Traditional Program

The traditional program (dispatch with losses) gives us the results shown in Table 4.

3.2 Associated Neural Network Program

Based on the results of the traditional dispatch program, we will create a neural network with P: input matrix [5x17] taken from Table 3, T: target matrix [3x17] taken from Table 4 representing the power generated by the three stations using traditional program. We will use two layers neural network with error backpropagation learning (Fig. 4):

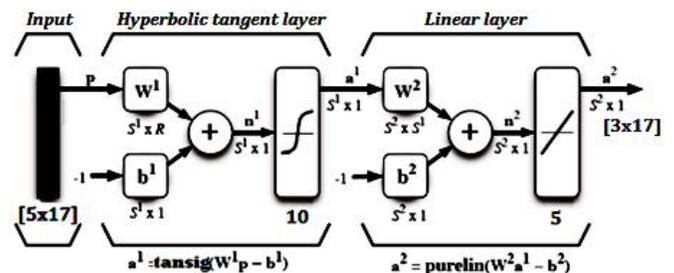


Fig. 4 Representation of the neural network used

The network contains: an input (P matrix), a hidden layer of 10 neurons with tansig (hyperbolic tangent sigmoid) activation

function and an output layer of 5 neurons with Pureline (linear) activation function. The learning results are shown in Table 4.

TABLE 4

RESULTS OF PLANTS GENERATIONS FOLLOWING TOTAL DEMANDS OF TABLE 3 USING TRADITIONAL AND NEURAL NETWORK PROGRAMS

TRADITIONAL PROGRAM RESULTS [MW] (target)			NEURAL NETWORK PROGRAM RESULTS [MW] (output)		
B1	B2	B3	B1	B2	B3
153.053	0	0	152.451	1.859	1.622
192.352	0	0	192.758	1.464	0.589
232.705	0	0	235.009	0.292	0.631
273.112	0	0	272.580	0.599	1.231
307.207	6.291	0	300.584	8.783	1.924
307.633	46.175	0	310.004	38.134	1.909
288.103	106.935	0	288.254	107.243	0.371
307.412	117.122	21.097	298.709	124.038	21.392
307.91	127.326	61.005	307.791	126.953	60.894
307.444	137.548	102.136	309.066	131.349	105.869
307.271	150.482	140.448	310.001	150.728	137.788
307.745	180.552	160.551	309.189	180.089	160.627
307.244	211.715	180.666	307.778	210.695	181.743
307.769	241.992	200.792	307.894	239.753	203.061
307.322	273.376	220.930	309.280	269.228	222.915
307.900	303.865	241.08	310.083	303.132	238.014
307.906	349.977	244.867	308.067	346.016	244.817

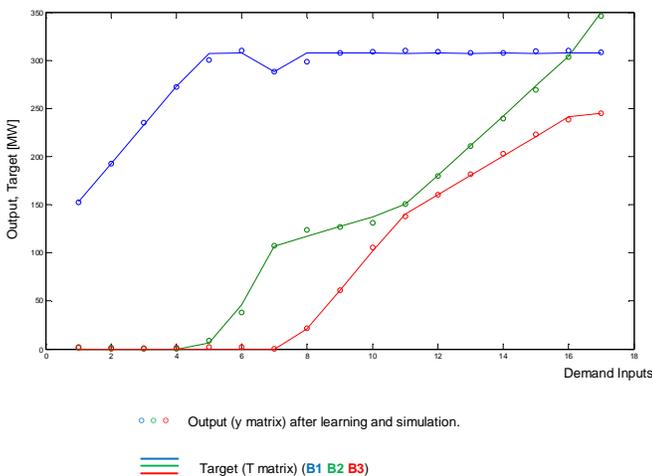


Fig. 5 Comparison between output (Y) and Target (T)

We can see from Table 4 and Fig. 5 that the values of the output are close to the target. We can deduce then that the training of our neural network is considered good. We must therefore test the network to judge the reliability of the learning.

3.3 Network Test

We introduce the matrix P_{test} which contains values in the borders of training matrix P, but did not participate in it.

TABLE 5

BUSBAR DEMANDS FOR NEURAL NETWORK TESTING

Ptest [MW]				
B4	B5	B6	B7	B8
0	75	29	44	119
0	77	31	46	121
11	117	71	86	161
25	131	85	100	175
53	159	113	128	203
65	171	125	140	215
72	178	132	147	222
86	192	146	161	236

then, P_{test} is passed through the traditional and Neural network programs. This will give us Ttest and Ytest respectively (Table 6).

TABLE 6

TEST RESULTS MATRICES TTEST AND YTEST

Ttest [MW]			Ytest [MW]		
B1	B2	B3	B1	B2	B3
269.069	0	0	269.784	0.001	0.004
277.156	0	0	276.264	0.971	0.003
307.460	118.142	25.079	309.767	116.283	24.360
307.172	132.435	82.044	307.092	128.125	86.656
307.892	189.592	166.584	307.629	189.400	167.042
307.503	226.841	190.727	307.038	228.499	189.488
307.878	248.060	204.819	306.984	250.808	202.875
307.666	291.657	233.018	307.694	290.522	234.198

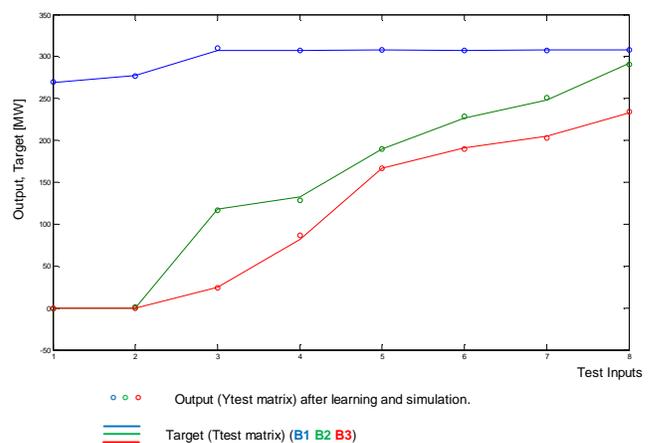


Fig. 6 Comparison between Ytest and Ttest

Based on data in Table 6 and Fig. 6, we can confirm that the learning of our neural network is considered good. To have a very good to excellent learning, output must be perfectly identical to the target ($Y \equiv T$), this requires:

- Using a good economic dispatch traditional program, based on precise data: power plants parameters, lines imped-

ances and grid topography;

- Simulating the largest number of cases based on actual statistics of busbars energy demands;
- Simulating random and unpredictable cases (fault cases) which are not included in the statistics;
- Reducing learning step of the neural network to make a significant interpolation for unknown data (unpredictable);
- Adding more data, i.e. increasing the size of P and T matrices, thus making a more detailed learning which covers the majority of cases from the traditional program (caution: very lengthy P and T matrices ask for more powerful processors otherwise we risk having slow program).

A good economic dispatch software based on neural networks, must have a complete and excellent learning and test, then, it could execute directly the dispatching, i.e. demand data could be presented in real time to the neural network program where the decisions take two cases: if the neural network finds data (real time) in its matrix P, it gives immediate release to predefined outputs from matrix T; if the neural network cannot find data in matrix P, it makes a direct interpolation to give results.

4 CONCLUSION

The most important difference between traditional and neural network programs is the execution time. The traditional program is slow because:

- It uses iterative loops : *if*, *while* and *for* ; which affects directly the execution time;
- It stocks a largest number of parameters: parameters of plants, lines and each busbar;
- The more a network is meshed the more the program is slow.

For the neural network program, the execution time is very rapid (milliseconds), because:

- It contains loops only on training; then after training the problem of dispatching becomes a classification problem.
- It is an executor; it performs data already stored and interpolates intermediate ones.

In practice, we opt for an economic dispatch with the fastest possible frequency (5 or 15 minutes instead of every hour [8]), even in real time [9], which is practically impossible with the traditional program.

To control power grids, time is very important and vital especially in economic dispatch and many other disciplines: protection, stability, power flow, etc. Using techniques of artificial intelligence and more specifically neural networks reduces the execution time; this will bring a huge economic gain by reducing losses, thus restraining the consumption of fuels (coal, oil, gas, uranium etc.). The decrease of production implies a contribution to the preservation of the environment by reducing pollution and global warming.

5 ACKNOWLEDGMENT

The authors gratefully acknowledge the contributions of Peter Kelen, research officer and owner of Power Optimization Software.

6 REFERENCES

- [1] ELEC234 : "Conduite des réseaux électriques", *Projet : dispatching économique*. Université libre de Bruxelles.
- [2] M.Younes, M.Rahli and M.Kandouci, "Répartition économique des puissances par un algorithme génétique en code réel "6ème Conférence francophone de modélisation et simulation - MOSIM'06 - du 3 au 5 avril 2006 - Rabat- Maroc.
- [3] A.Merev, "Comparison of the economic dispatch solutions with and without transmission losses" *Istanbul university engineering faculty. JEE*, Vol.2, No. 2, 2002.
- [4] J-F. Jodouin, *Les réseaux de neurones : principe et définition*. Lavoisier, Septembre 1994.
- [5] M. Parizeau, *Réseau de neurones GIF-21140 et GIF -64326*. Université LAVAL, Automne 2004.
- [6] P. Borne, M.Benrejeb and J.Haggège, *Les réseaux de neurones : présentation et application*. TECHNIP, 2007.
- [7] C-W.Chua, *A stochastic pool-based electricity market simulator*. McGill University, Montreal, November 2000.
- [8] "Economic Dispatch: Concepts, Practices and Issues" *FERC Staff*, California, November 13, 2005.
- [9] Subject: "Security Constrained Unit Commitment and Real-Time Commitment Rules" TECHNICAL BULLETIN 51, 11/10/2004; revised 9/9/09.

Review on Twisted Tapes Heat Transfer Enhancement

C.Nithiyesh Kumar, P.Murugesan

Abstract— Heat transfer augmentation techniques refer to different method used to increase rate of heat transfer without affecting much the overall performance of the system. These techniques are used in heat exchangers. Some of the applications of heat exchangers are in process industries, thermal power plant, air conditioning equipment, refrigerators, radars for space vehicles, automobiles etc. In the past decade, several studies on the passive techniques of heat transfer augmentation have reported. The present paper review mainly focus on the twisted tape heat transfer enhancement and its design modification towards the enhancement of heat transfer and saving pumping power.

Index Terms— Heat transfer augmentation, Passive methods, Tape inserts, Reynolds number, Friction factor

1 INTRODUCTION

Heat transfer enhancement or augmentation techniques refer to the improvement of thermo hydraulic performance of heat exchangers. Existing enhancement techniques can be broadly classified into three different categories:

1.1 Active Techniques: These techniques are more complex from the use and design point of view as the method requires some external power input to cause the desired flow modification and improvement in the rate of heat transfer. It finds limited application because of the need of external power in many practical applications.

1.2 Passive Techniques: These techniques do not require any direct input of external power; rather they use it from the system itself which ultimately leads to an increase in fluid pressure drop. They generally use surface or geometrical modifications to the flow channel by incorporating inserts or additional devices. They promote higher heat transfer coefficients by disturbing or altering the existing flow behavior except for extended surfaces.

1.3 Compound Techniques: When any two or more of these techniques are employed simultaneously to obtain enhancement in heat transfer that is greater than that produced by either of them when used individually, is termed as compound enhancement¹

In this paper, a review of heat transfer enhancement using twisted tape and its modification is done. This paper also gives the performance criteria of different twisted tape inserts. Finally it is expected to be the pioneer source as an intensive literature review for twisted tape swirl generator.

Abbreviation

F	Friction factor
N_u	Nusselt number
R_e	Reynolds number
S_w	Swirl parameter
S	Spacing between two twisted tapes
TT	Twisted tape
D_i	Inside diameter
γ	Twist ratio
HTE	Heat transfer enhancement
α	Twist angle
g	Thermal performance factor
Pr	Prandtl number
η	Thermo hydraulic efficiency
G	Mass velocity
TA	Twisted tape with alternate axis
CCTA	Clockwise and counter clockwise TA
VTT	V-Cut twisted tape
WR	Width ratio
DR	Depth ratio
STT	Serrated twisted tape
CCT	Centre cleared twisted tape
CTs/CoTs	Twin-counter/co-twisted tapes
ETT	Edgefold-Twisted Tape
WTA	Centre wing and alternate axis Twisted tape
P-TA	Peripheral cut- Alternate axis Twisted tape
O/S DWT	Oblique/ Straight - Delta winglet Twisted tape
D/I coil	Decreasing/increasing coil pitch ratio arrangement
BTT	Broken twisted tape

• PG Scholar, Department of Mechanical Engineering, K.S.Rangasamy College of Technology, Tiruchengode, Namakkal District, Tamilnadu, India-637215
Email ID: nithiyesh@gmail.com
• Department of Mechanical Engineering, K.S.R Institute for Engineering and Technology, Tiruchengode, Namakkal District, Tamilnadu, India-637215

2 TERMINOLOGY USED IN TWISTED TAPE

2.1 Thermo Hydraulic Performance For a particular Reynolds number, the thermo hydraulic performance of an insert is said to be good if the heat transfer coefficient increases significantly with a minimum increase in friction factor. Thermo hydraulic performance estimation is generally used to compare the performance of different inserts under a particular fluid flow condition.

2.2 Overall Enhancement Ratio The overall enhancement ratio is defined as the ratio of the heat transfer enhancement ratio to the friction factor ratio.

2.3 Nusselt Number The Nusselt number is a measure of the convective heat transfer occurring at the surface and is defined as hd/k , where h is the convective heat transfer coefficient, d is the diameter of the tube and k is the thermal conductivity.

2.4 Prandtl Number The Prandtl number is defined as the ratio of the molecular diffusivity of momentum to the molecular diffusivity of heat.

2.5 Pitch The Pitch is defined as the distance between two points that are on the same plane, measured parallel to the axis of a Twisted Tape.

2.6 Twist Ratio The twist ratio is defined as the ratio of pitch length to inside diameter of the tube.

3. REVIEW ON TWISTED TAPE

The present paper contributes for review of twisted tape inserts. The main objective of this paper is to review the work carried on plain twisted tape, modified twisted tape, and modified twisted tape geometry.

3.1 Plain Twisted Tape

Behabadi et al. [1] experimental investigated the heat transfer coefficients and pressure drop during condensation of HFC-134a in a horizontal tube fitted with TT. The refrigerant flows in the inner copper and the cooling water flows in annulus. Also empirical correlations were developed to predict smooth tube and swirl flow pressure drop.

Syam Sundar and sharma [2] investigated the thermo physical properties like thermal conductivity and viscosity of Al_2O_3 nanofluid is determined through experiments at different volume concentrations and temperatures. From the result it is observed that, heat transfer coefficients and ' f ' is higher when compared to water in a plain tube. Also a generalized regression equation is developed with the experimental data for the estimation of ' f ' and ' Nu '.

Promvonge et al. [3] experimentally investigated the heat transfer rate, ' f ' and ' η ' of the combined devices of TT and wire coil. The experiment is carried out by arranging in two different forms: (1) D-coil and (2) DI-coil while the TT was prepared with two different twist ratios.

Klaczak [4] investigated experimentally the heat transfer for laminar flow of water in an air cooled vertical copper pipe with TT inserts of various pitch value. The tests were executed for laminar flow within $110 \leq Re \leq 1500$, $8.1 \leq Gz \leq 82.0$ and $1.62 \leq y \leq 5.29$. Result shows that the heat transfer increases with increase in Twisted tape pitch value.

3.2 Modified Twisted Tape

Ferroni et al. [5] experimentally analyzed, the isothermal pressure drop tests, were performed on horizontal round tube with equally spaced and short-length TT. Various test are made with $1.5 \leq y \leq 6$ and $30 \leq S \leq 50$. The Darcy friction factor associated with the tested ' y ' and ' S ' combinations was calculated, and a relation correlating this factor to ' Re ', ' y ' and ' S ' was developed.

Changhong Chen et al. [6] analyzed the computational fluid dynamics (CFD) modeling for the optimization of regularly spaced short-length TT in a circular tube. The configuration parameters are given by the ' S ', ' y ' and ' α '. The result is made such that the mean heat transfer and flow resistance increase with an increase in α .

Yadav [7] experimentally investigated on the half length TT insertion on heat transfer & pressure drop characteristics in a U-bend double pipe heat exchanger. The experimental results revealed that the increase in heat transfer rate of the TT inserts is found to be strongly influenced by tape-induced swirl.

Eiamsa-ard et al. [8] made a comparative investigation of enhanced heat transfer and pressure loss by insertion of single TT, full-length dual TT and regularly-spaced dual TT as swirl generators. The result shows that all dual TT with free spacing yield lower heat transfer enhancement in comparison with the full-length dual TT.

Hata and masuzakib [9] investigated the TT- induced swirl flow heat transfer due to exponentially increasing heat inputs with various exponential periods and the TT-induced pressure drop were systematically measured. The influence of ' y ' and ' Re ' based on swirl velocity, ' Re_{sw} ' on the TT-induced swirl flow heat transfer was investigated and predictable correlation was derived.

Eiamsa-ard et al. [10] studied the influences of multiple twisted tape vortex generators (MT-VG) on the heat transfer and fluid friction characteristics in a rectangular channel. From the experiment it is revealed that, the channel with the ' y ' and ' S ' provides higher heat transfer rate and pressure loss than those with the larger ' y ' and free-spacing ratio under similar operation condition.

Eiamsa-ard et al. [11] an experimental study on the mean ' Nu '; ' f ' and ' g ' in a round tube with short-length TT insert. The full-length twisted tape is inserted into the tested tube at a single $y = 4.0$ while the short-length tapes mounted at the entry test section. The experimental result indicates that the

presence of the tube with short-length twisted tape insert yields higher heat transfer rate.

Eiamsa-ard et al. [12] mathematically investigated the swirl flow in a tube induced by loose-fit twisted tape insertion. Effects of the clearance ratio on 'Nu', 'f' and 'g' are numerically investigated for TT at two different twist ratios.

Thianpong et al. [13] investigated experimentally the friction and compound heat transfer behaviors in a dimpled tube fitted with a TT swirl generator, using air as working fluid. The experiments are conducted by using two dimpled tubes with different pitch ratios and three twisted tapes with three different twist ratios. It is reveal that both heat transfer coefficient and 'f' in the dimpled tube fitted with the TT, are higher than those in the dimple tube acting alone and plain tube.

Promvongse and Eiamsa-ard [14] investigated thermal characteristics in a circular tube fitted with conical-ring and a TT swirl generator. The experimental results reveal that the tube fitted with the conical-ring and TT provides 'Nu' values of around 4 to 10% and enhancement efficiency of 4 to 8% higher than that with the conical-ring alone.

Mengna et al. [15] investigated experimentally the Pressure drop and compound heat transfer characteristics of a converging-diverging tube with evenly spaced TT (CD-T tube). Swirl was generated by evenly spaced twisted-tape elements which vary in twist ratio and rotation angle.

Eiamsa-ard et al. [16] experimentally investigated on the 'HTE' and 'f' characteristics in a double pipe heat exchanger fitted with regularly TT insert. By comparing the result with plain tube, it is evident that the heat transfer coefficient increased with 'y' and 'S'.

Saha et al. [17] experimentally investigated the HTE and pressure drop characteristics in the tube with regularly spaced TT element. From the result, it is observed that 'Pinching' of tape rather than in connecting the tape element with rods is better proposition from thermohydraulic point of view.

3.3 Modified Geometry Twisted Tape

Wei Liu et al. [18] investigated numerically the HTE and 'f' characteristics of laminar flow in a tube with short-width and CCT. It is given that CCT is good technique in lamina flow and the heat transfer can be enhanced with a change in central clearance ratio.

Eiamsa-ard and wongcharee [19] experimentally investigated HTE, 'f' and 'g' characteristics of CuO/water nanofluid and modified 'TA'. The use of nanofluid with the TA provides considerably higher 'Nu' and 'g' than that of nanofluid with the PTT.

Eiamsa-ard et al. [20] studied the effect of 'Nu', 'f' and 'g' behaviors of tubes fitted with C-CTA. The results reveal that, 'Nu', 'f' and 'g' associated by TA are higher than those associated by PTT.

Murugesan et al. [21] investigated experimentally the 'HTE', 'f' and 'g' characteristics of tube fitted with VTT. The obtained results show that the mean Nusselt number and the mean 'f' in the tube with 'VTT' increases with in decrease 'y'.

Eiamsa-ard et al. [22] experimentally investigated the influences on 'Nu', 'f' and 'g' of CT/CoT tapes fitted in tube. The (CTs) are used as counter-swirl flow generators while (CoTs) are used as co-swirl flow generators. The results also show that the CTs are more efficient than the CoTs for HTE.

Eiamsa-ard et al. [23] present an experimental study of turbulent heat transfer and flow friction characteristics in a circular tube equipped with C-CC TT. The results shows that the HTE of the C-CC TT increases with the decrease of twist ratio and the increase of twist angle values.

Zhang and Mao [24] carried out the 3D numerical and experimental study of the heat transfer characteristics and the pressure drop of air flow in a circular tube with ETT and STT inserts. From the experimental study it is found that the 'η' slowly decreases as the 'y' and 'S' increases.

Eiamsa-ard et al. [25] presented an investigation of the effect of twisted tape with serrated edge insert. The use of the STT leads to higher heat transfer rate and friction factor than that of the TT for all cases. The 'g' of the STT tube under constant pumping power is above unity.

Saha [26] experimentally studied the heat transfer and the pressure drop characteristics of rectangular and square ducts with TT insert with oblique teeth. From experiment it is found that, the axial corrugation in combination with TT with oblique teeth performs better than those without oblique teeth.

Eiamsa-ard et al. [27] experimentally studied the effects of the twisted tapes consisting of WT-A in a tube. It is found that, 'Nu', 'f', 'g' provided by the WT-A is higher than other type of tapes.

Eiamsa-ard and seemawute [28] experimentally investigated the effect of PT-A on the fluid flow and HTE characteristic. From the result, it is revealed that the PT-A offer the maximum thermal performances at constant pumping power.

Eiamsa-ard et al. [29] investigated 'HTE', 'f' and 'g' characteristics in a tube fitted with DWT. Influences of the O-DWT and S-DWT arrangements are also described. The obtained results show that the thermal performance factor in the tube with O-DWT is greater than that with S-DWT.

Eiamsa-ard et al. [30] Investigated the Effects of PTT insert on heat transfer, 'f' and 'g' characteristics in a round tube. Nine different PTT with 'y', different 'DR' and different 'WR' were tested. From the result, it is revealed that 'Nu', 'f' and 'g' are found to be increased with 'DR' and 'WR'.

Radhakrishnan. et al. [31] made experimental investigation on 'HTE', 'f' and 'g' of thermosyphon solar water heater system fitted with full-length twist, twist fitted with rod and spacer fitted at the trailing edge. Conclusions made from the results show that 'HTE' in TT collector is higher than the plain tube.

Bharatdwaj et al. [32] experimentally determined pressure drop and heat transfer characteristics of flow of water in a 75 start spirally grooved tube with twisted tape insert are presented. It is found 'HTE' in spiral tube is higher when compared to plain tube.

Chang et al. [33] experimental study that comparatively examined the spiky twisted-tape insert (swirl tube) placed in

a tube. The dispersed rising air bubbles in the plain tube and the centrifugal-force induced coherent spiral stream of coalesced bubbles in the swirl-tube core considerably modify the pressure-drop and heat-transfer performances from the single-phase conditions.

Eiamsa-ard et al. [34] experimentally investigated the 'HTE' and 'f' effect in C-CC TT in heat exchanger. The experimental result revealed that heat transfer rate and 'f' is high compared to PTT.

Chang et al. [35] experimentally examined the turbulent heat transfer in a swinging tube with a STT insert under seagoing conditions. This swirl tube swings about two orthogonal axes under single and compound rolling and pitching oscillations. Synergistic effects of compound rolling and pitching oscillations with either harmonic or non-harmonic rhythms improve heat transfer performances.

Murugesan. et al. [36] experimentally investigated the heat transfer and 'f' characteristics of trapezoidal -cut TT with $y=$

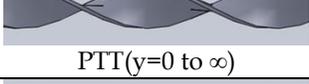
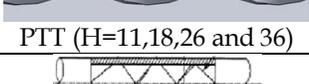
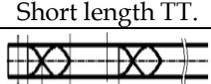
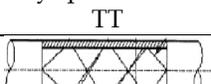
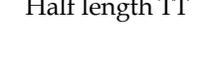
4.0 and 6.0. From the experiment it is revealed, that there was a significant increase in heat transfer coefficient and 'f' for tape with trapezoidal-cut.

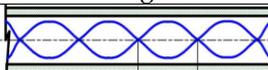
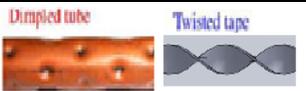
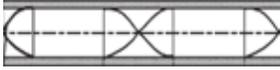
Chang et al. [37] Studied experimentally on compound heat transfer enhancement in a tube fitted with STT. The serrations on two sides of the TT with twist ratio ' y ' = 1.56, 1.88, 2.81 ∞ are the square-sectioned ribs with the identical rib-pitch and rib-height. From the experiment it is revealed that the friction factor and heat transfer rate is comparatively high than PTT.

Chang et al. [38] made an experimental study in measuring the axial heat transfer distributions and the pressure drop coefficients of the tube fitted with BTT. From the experimental result it is revealed that local 'Nu' and mean 'f' in the tube fitted with the BTT increase as the ' y ' decreases.

The summary of the above work is given in the table 1.0 based on the classification of the tape, type of flow, observation and comment.

Table 1.0 Summary of Important Investigation on Twisted Tape

S.No	AUTHOR NAME	TYPE OF FLOW	CONFIGURATION OF TAPE	OBSERVATION	COMMENT
1.	Behabadi et al.	Turbulent	 PTT ($y=6, 9, 12$ and 15)	Decrease in ' y ' increase the heat transfer rate. Pressure drop increases with reduction in ' y '	It is observed that the TT with ' y ' = 9 gives the best ' g ' with minimum pressure drop.
2.	Syam Sundar & sharma	Turbulent	 PTT ($y=0 < H/D < 83$)	No increase in pressure drop or 'f' when compared to water.	The maximum ' g ' in the result is obtained at $y=5$.
3.	Promvonge et al.	Turbulent	 PTT($y=0$ to ∞)	The heat transfer and 'f' is more efficient in TT than wire coil.	At low Re, with $y=3$ the thermal performance is higher.
4.	Klaczak	Laminar	 PTT ($H=11,18,26$ and 36)	The ratio Nu_T/Nu has the highest value when 'H' was the smallest.	The heat transfer coefficient increases with increase in twist tape pitch value
5.	Ferroni et al.	Turbulent	 Short length TT.	'f' and 'Nu' is low for short length TT and the performance is good.	The multiple short length twist tape performance is high compared to other.
6.	Chang et al.	Turbulent	 Regularly spaced short length TT	Heat transfer and flow resistance increase with an increase in the ' α '.	The larger ' α ' yields a higher heat transfer value and a greater flow resistance.
7.	Yadav	Laminar	 Half length TT	On unit pressure drop basis and on unit pumping power basis, half length TT is more efficient than full-length TT.	On unit pressure drop basis, the heat transfer performance of smooth tube is maximum followed by half-length TT.

8.	Eiamsa-ard et al.	Turbulent	 Regularly spaced dual TT	The 'f' from the dual TT increases up to 23% over the single twisted tape.	The 'f' tends to decrease with the rise of 'Re' and 'y'.
9.	Hata & Masuzakib	Turbulent	 PTT ($y = 2.39-4.45$)	The values of ΔP become linearly higher with an increase in the Re_d , and also influence of 'y' and 'Re' based on swirl velocity, Re_{sw}	Overall enhancement ratio increases with tighter twist ratio and decreases with increase in Reynolds number.
10.	Eiamsa-ard et al.	Turbulent	 Multiple TT	'Nu' increases in the range of 170% and 'f' in the range of 1.45 and 5.7. Heat transfer rate is 1.4 than the plain tube.	The channel with the smaller 'y' and 'S' provides higher heat transfer rate and pressure loss.
11.	Eiamsa-ard et al.	Turbulent	 Short length TT	The value of f and heat transfer rate ranges from 1.76-1.99, 1.16- 1.27.	Short length TT is better than full length TT on basis of 'g'
12.	Eiamsa-ard et al.	Turbulent	 Loose fit TT	Nu/Nu0 decreases with increase in Re.	The TT with tight fit gives high heat transfer rate but decrease in the 'f'.
13.	Thianpong et al.	Turbulent	 Dimpled tube fitted with TT	Heat transfer and 'f' increases with 'y' decreases.	Dimpled tube with TT gives higher heat transfer coefficient than TT.
14.	Promvongse & Eiamsa-ard	Turbulent	 Conical ring with twisted tape	With $y=3.75$ the ' η ' is 1.96.	Combined device increases the thermal performance.
15.	Mengna et al.	Turbulent	 TT in CD tube	At $Y=4.762$ and $\theta=180^\circ$ the best heat transfer and friction factor is obtained.	Converging diverging (CD) tube with TT creates swirling motion to fluid and increase the time of contact.
16.	Eiamsa-ard et al.	Turbulent	 Regularly spaced twisted tape	Heat transfer coefficient and friction factor increases with increase in space ratio.	
17.	Saha et al.	Laminar	 Regularly spaced twisted tape	Pinching of tape rather than connecting tape element give high 'g'.	Uniform pitch performs better than gradually decreasing pitch.
18.	Wei liu et al.	Laminar	 Clear centered twisted tape	CCT gives 'g' value 7-20% more than TT.	CCT gives more fluid flow and optimal 'f'.
19.	Eiamsa-ard & Wongcharee	Laminar	 Alternate axis Twisted tape	'Nu' increases with increase in 'Re' and fluid concentration.	TA makes more swirl in fluid flow with increased efficiency.
20.	Eiamsa-ard et al.	Laminar	 Clockwise and counter clockwise TA	At $y=3$ and $Re=830$, more efficient heat transfer is obtained.	Friction factor in TA increases with decrease in twist ratio.

21.	Murugesan et al.	Turbulent	 V-cut Twisted tape	Influence of 'DR' was more dominant than 'WR' for all 'Re'.	V cut TT gives higher transfer rate and friction factor than plain tube.
22.	Eiamsa-ard et al.	Turbulent	 Twin-counter/co-twisted tapes	'Nu', 'f' and 'g' increases with decrease in 'y'.	Heat transfer rate in tube with CTs are higher than those with CoTs.
23.	Eiamsa-ard et al.	Turbulent	 Clockwise and counter clockwise TA	Heat transfer rate of the CCTA increases with decrease of twist ratio.	Counter clockwise cut in twisted tape creates more of fluid inside the tube
24.	Zhang & Mao	Turbulent	 Edgefold Twisted tape	The highest performance of this TT is 140% when gap width reduces to 1 mm.	ETT with minimum α gives maximum thermal performance.
25.	Eiamsa-ard et al.	Turbulent	 Serated Twisted tape	Heat transfer rate is 1.77 times higher than that of plain tube.	STT increases turbulence intensity and breaking down the boundary layer.
26.	Saha	Turbulent	 Oblique teeth Twisted tape	At constant pumping power, the heat transfer rate is 1.55	TT with oblique teeth show high performance than TT without Oblique teeth.
27.	Eiamsa-ard et al.	Turbulent	 Centre wing and alternate axis Twisted tape.	Nu, f and g are higher in WTA than plain TT.	WTA will create the effect of swirling flow, and strong collision of the streams.
28.	Eiamsa-ard & Seemawute	Turbulent	 Peripheral cut- Alternate axis Twisted tape	Heat transfer rate by using PTA is 184%.	PTA provides maximum thermal performance at constant pumping power.
29.	Eiamsa-ard et al.	Turbulent	 DWT	'Nu' and 'f' increases with decrease in 'Re'.	DWT can replace any of the TT to reduce size of the heat exchanger.
30.	Eiamsa-ard et al.	Laminar	 PTA	'Nu', 'f' and 'g' increased with increase in depth ratio.	PTA provides high 'g' with constant pumping power
31.	Radha Krishnan et al.	Laminar	 TT with rod and spacer	'Nu' is 13.5% higher than PTT and 'f' is given as 14.85.	This is well effective for laminar flow only.
32.	Bharatdwaj et al.	Laminar	 Spirally grooved tube with TT	Heat transfer enhancement is increased due to swirl flow.	Spirally grooved tube will be well effective only with TT inside the tube.
33.	Chang et al.	Turbulent	 Spiky Twisted tape	HTE and 'g' of spiky TT is higher than that of PTT.	
34.	Eiamsa-ard et al.	Turbulent	 CCTA	Heat transfer rate is 219% of the plain tube.	

35.	Chang et al.	Turbulent	 Serrated Twisted tape	Synergic effect of rolling and pitching oscillations improve heat transfer rate.	Serrated tube cause swirl motion and oscillation to the flow and increases 'g'.
36.	Murugesan et al.	Turbulent	 Trapezoidal cut TT	By the trapezoidal cut TT, the 'f' and 'g' increases with decrease in twist ratio.	Trapezoidal cut TT, increase in HTE value with constant pumping power.
37.	Chang et al.	Turbulent	 Serrated Twisted tape	Nu and f increases as the twist ratio decreases.	Serration in the tube cause swirling in the boundary layer of the tube.
38.	Chang et al.	Turbulent	 Broken twisted tape	By using broken TT, the $f=2-4.7$ and $g=1.8$.	Broken TT increase the fanning factor with decrease in twist ratio.

CONCLUSION

This review has considered heat transfer and pressure drop investigations of the various twisted tape placed in heat exchangers. Almost all possible research subjects have been summarized on the case in the literature, such as heat transfer and pressure drop studies according to plain twisted tape, modified twisted tape, and modified twisted tape geometry.

A twisted tape and modified twisted tape inserts mixes the bulk flow well and therefore performs better in laminar flow, because in laminar flow the thermal resistant is not limited to a thin region. The result also shows twisted tape insert is more effective in laminar flow, and pressure drop penalty is created during turbulent flow.

In case of twisted tape with modified geometry, more turbulence is created during the swirl of fluid and gives higher heat transfer rate compared to plain twisted tape and modified twisted tape. The result shows that for modified twisted tape geometry, the heat transfer rate is higher with reasonable friction factor for both laminar and turbulent flow.

These conclusions are very useful for the application of heat transfer enhancement in heat exchanger networks.

REFERENCE

- [1] V. Hejazi, M.A. Akhavan-Behabadi and A. Afshari, Experimental investigation of twisted tape inserts performance on condensation heat transfer enhancement and pressure drop, *International Communications in Heat and Mass Transfer*, Vol. 37 (2010), pp. 1376-1387.
- [2] L. Syam Sundar and K.V. Sharma, Turbulent heat transfer and friction factor of Al_2O_3 Nanofluid in circular tube with twisted tape inserts, *International Journal of Heat and Mass Transfer*, Vol. 53 (2010), pp. 1409-1416.
- [3] S. Eiamsa-ard, P. Nivesrangsan, S. Chokphoemphun and P. Promvonge, Influence of combined non-uniform wire coil and twisted tape inserts on thermal performance characteristic, *International Communications in Heat transfer*, Vol. 37 (2010), pp. 850-856.
- [4] A. Klaczak, Heat transfer by laminar flow in a vertical pipe with twisted-tape inserts, *Heat and Mass Transfer*, Vol. 36 (2000), pp. 195-199, Springer-Verlag 2000.
- [5] P. Ferroni, R.E. Block, N.E. Todreas and A.E. Bergles, Experimental evaluation of pressure drop in round tubes provided with physically separated, multiple, short-length twisted tapes, *Experimental Thermal and Fluid Science* xxx (2011) xxx-xxx.
- [6] Yangjun Wang, Meiling Hou, Xianhe Deng, Li Li, Cheng Huang, Haiying Huang, Gangfeng Zhang, Changhong Chen and Weijun Huang, Configuration optimization of regularly spaced short-length twisted tape in a circular tube to enhance turbulent heat transfer using CFD modeling, *Applied Thermal Engineering*, Vol. 31 (2011), pp. 1141-1149.
- [7] Anil Yadav, Effect of half length twisted tape turbulator on heat transfer & pressure drop characteristics inside a double pipe U-bend heat exchanger, *Jordan journal of Mech. & Industrial engg.*, Vol. 3 No.1 (2009), pp. 17-22.
- [8] S. Eiamsa-ard, Chinarak Thianpong, Petpices Eiamsa-ard and Pongjet Promvonge, Thermal characteristics in a heat exchanger tube fitted with dual twisted tape elements in tandem, *International Communications in Heat and Mass Transfer* Vol. 37 (2010), PP.39-46.
- [9] K. Hataa and S.Masuzakib, Twisted-tape-induced swirl flow heat transfer and pressure drop in a short circular tube under velocities controlled, *Nuclear Engineering and Design*, xxx (2010) xxx-xxx, NED-5980.
- [10] S. Eiamsa-ard, Study on thermal and fluid flow characteristics in turbulent channel flows with multiple twisted tape vortex generators, *International Communications in Heat and Mass Transfer*, Vol. 31 (2010), pp. 644-651.
- [11] S. Eiamsa-ard, Chinarak Thianpong, Petpices Eiamsa-ard and Pongjet Promvonge, Convective heat transfer in a circular tube with short-length twisted tape insert, *International communication in Heat and Mass Transfer*, Vol. 36 (2009), pp. 365-371.
- [12] S. Eiamsa-ard, K. Wongcharee and S. Sripattanapipat, 3-D Numerical simulation of swirling flow and convective

- heat transfer in a circular tube induced by means of loose-fit twisted tapes, *International communication in Heat and Mass Transfer* Vol. 36 (2009), pp. 947-955.
- [13] Chinaruk Thianpong, Petpices Eiamsa-ard, Khwanchit Wongcharee and Smith Eiamsa-ard, Compound heat transfer enhancement of a dimpled tube with a twisted tape swirl generator, *International Communications in Heat and Mass Transfer* Vol. 36 (2009), pp. 698-704.
- [14] P. Promvong and S. Eiamsa-ard, Heat transfer behaviors in a tube with combined conical ring and twisted-tape inserts, *International Communications in Heat and Mass Transfer* Vol. 34 (2007), pp. 849-859.
- [15] HONG Mengna, DENG Xianhe, HUANG Kuo and LI Zhiwu, Compound heat transfer enhancement of a converging-diverging tube with evenly spaced twisted-tapes, *Chinese Journal of Chemical Engineering*, Vol. 15 No.6 (2007), pp. 814-820.
- [16] Smith Eiamsa-ard, Chinaruk Thianpong and Pongjet Promvong, Experimental investigation of heat transfer and flow friction in a circular tube fitted with regularly spaced twisted tape elements, *International Communications in Heat and Mass Transfer* Vol. 33 (2006), pp. 1225-1233.
- [17] Saha, S. K, Dutta, A. and Dhal, S. K., Friction and heat transfer characteristics of laminar swirl flow through a circular tube fitted with regularly spaced twisted-tape elements, *International Journal of Heat and Mass Transfer*, Vol. 44 (2001), pp. 4211-4223.
- [18] Jian Guo, Aiwu Fan, Xiaoyu Zhang and Wei Liu, A numerical study on heat transfer and friction factor characteristics of laminar flow in a circular tube fitted with center-cleared twisted tape, *International Journal of Thermal Sciences* Vol. 50 (2011), pp. 1263-1270.
- [19] K.Wongcharee and S. Eiamsa-ard, Enhancement of heat transfer using CuO/water nanofluid and twisted tape with alternate axis, *International Communications in Heat and Mass Transfer* Vol. 38 (2011), pp. 742-748.
- [20] P. Murugesan, K. Mayilsamy, S. Suresh and P.S.S. Srinivasan, Heat transfer and pressure drop characteristics in a circular tube fitted with and without V-cut twisted tape insert, *International Communications in Heat and Mass Transfer* Vol. 38 (2011), pp. 329-334.
- [21] S. Eiamsa-ard, C. Thianpong and P. Eiamsa-ard, Turbulent heat transfer enhancement by counter/co-swirling flow in a tube fitted with twin twisted tapes, *Experimental Thermal and Fluid Science* Vol. 34 (2010), pp. 53-62.
- [22] Smith Eiamsa-ard and Pongjet Promvong, Performance assessment in a heat exchanger tube with alternate clockwise and counter-clockwise twisted-tape inserts, *International Journal of Heat and Mass Transfer* Vol. 53 (2010), pp. 1364-1372.
- [23] CUI Yong-zhang and TIAN Mao-cheng, Three-Dimensional Numerical Simulation of Thermal hydraulic Performance of a circular tube with Edgefold-Twisted-Tape Inserts, *Journal Of Hydrodynamics*, Vol. 22 No. 5 (2010), pp. 662-670.
- [24] Smith Eiamsa-ard and Pongjet Promvong, Thermal characteristics in round tube fitted with serrated twisted tape, *Applied Thermal Engineering* Vol. 30 (2010), pp. 1673-1682.
- [25] S.K. Saha, Thermohydraulics of turbulent flow through rectangular and square ducts with axial corrugation roughness and twisted-tapes with and without oblique teeth, *Experimental Thermal and Fluid Science* Vol. 34 (2010), pp. 744-752.
- [26] S. Eiamsa-ard, K. Wongcharee, P. Eiamsa-ard and C. Thianpong, Thermohydraulic investigation of turbulent flow through a round tube equipped with twisted tapes consisting of centre wings and alternate-axes, *Experimental Thermal and Fluid Science* Vol. 34 (2010), pp. 1151-1161.
- [27] P. Seemawute and S. Eiamsa-ard, Thermohydraulics of turbulent flow through a round tube by a peripherally-cut twisted tape with an alternate axis, *International Communications in Heat and Mass Transfer*, Vol. 37 (2010), pp. 652-659.
- [28] S. Eiamsa-ard, K. Wongcharee, P. Eiamsa-ard and C. Thianpong, Heat transfer enhancement in a tube using delta-winglet twisted tape inserts, *Applied Thermal Engineering*, Vol. 30 (2010), pp. 310-318.
- [29] S. Eiamsa-ard, Panida Seemawute and K. Wongcharee, Influences of peripherally-cut twisted tape insert on heat transfer and thermal performance characteristics in laminar and turbulent tube flows, *Experimental Thermal and Fluid Science*, Vol. 34 (2010), pp. 711-719.
- [30] S. Jaisankar, T.K. Radhakrishnan and K.N. Sheeba, Experimental studies on heat transfer and friction factor characteristics of thermosyphon solar water heater system fitted with spacer at the trailing edge of twisted tapes, *Applied Thermal Engineering* Vol. 29 (2009), pp. 1224-1231.
- [31] Bharadwaj, A.D. Khondge and A.W.Date, Heat transfer & pressure drop in spirally grooved tube with twisted tape insert, *J. Heat Transfer*, Vol. 52 No.5 (2009), PP. 1938-1944.
- [32] S.W. Chang, Arthur William Lees and Hsien-Tsung Chang, Influence of spiky twisted tape insert on thermal fluid performances of tubular air-water bubbly flow, *International Journal of Thermal Sciences* Vol. 48 (2009), pp. 2341-2354.
- [33] B. Silapakijwongkul, S. Eiamsa-ard and P. Promvong, Effect of clockwise and counterclockwise twisted-tape turbulators on heat transfer augmentation inside a double pipe heat exchanger. *International Energy Journal*.
- [34] S.W. Chang, Tong-Miin Liou, Jin Shuen Liou and Kun-Tse Chen, Turbulent heat transfer in a tube fitted with serrated twist tape under rolling and pitching environments with applications to shipping, *Journal of Ocean Engineering*, Vol. 35 (2008), pp. 1569-1577.
- [35] P. Murugesan, K. Mayilsamy, S. Suresh and P.S.S. Srinivasan, Heat Transfer And Pressure Drop Characteristics of Turbulent Flow in a Tube Fitted With

- Trapezoidal-Cut Twisted Tape Insert, International Journal Of Academic Research, Vol. 1. No. 1 (2009), PP. 123-128
- [37] S.W. Chang, Yih Jena Jan and Jin Shuen Liou, Turbulent heat transfer and pressure drop in tube fitted with serrated twisted tape, International Journal of Thermal Sciences Vol. 46 (2007), pp. 506-518.
- [38] S.W Chang, Tsun Lirng Yang and Jin Shuen Liou, Heat transfer and pressure drop in tube with broken twisted tape insert, Experimental Thermal and Fluid Science Vol. 32 (2007), pp. 489- 501.

Attrition Issues and Retention Challenges of Employees

Brijesh Kishore Goswami, Sushmita Jha

Abstract— In today's fiercely competitive business landscape, success hinges much on the retention of employees. In a knowledge driven economy, it is the people who are emerging as key competitive differentiators and retaining the talent has become a matter of paramount importance. But swelling attrition levels across the businesses are strait-jacketing the distraught HR practitioner's ability to come out with good retention strategies. Willy-nilly, attrition is here to stay and is not a fringe concern for the companies any more. The trick lies in having sound human resource practices in place that can effectively combat the scourge of attrition. This article would provide an insight into the various issues surrounding the problem of attrition and what major challenges are being faced in retaining the employees. Acquiring skilled workforce is not just enough in today's changing economy; instead a lot needs to be done to retain them. Losing knowledgeable and trained employees can cause serious damage to the company's progress and performance in the market.

Index Terms—

1 INTRODUCTION

Companies in India as well as in other countries face a formidable challenge of recruiting and retaining talents while at the same time having to manage talent loss through attrition be that due to industry downturns or through voluntary individual turnover. Losing talents and employees result in performance losses which can have long term negative effect on companies especially if the departing talent leaves gaps in its execution capability and human resource functioning which not only includes lost productivity but also possibly loss of work team harmony and social goodwill. With attrition rates being a bane of every industry, companies are devising innovative business models for effective retention of talent. There are a lot of factors responsible for attrition and employers are getting increasingly conscious of the factors that can keep an employee committed.

2 ATTRITION ISSUES

Attrition may be defined as gradual reduction in membership or personnel as through retirement, resignation or death. In other words, attrition can be defined as the number of employees leaving the organization which includes both voluntary and involuntary separation. The employee gradually reduces his/her ties with the company than crib about the underlying factors causing attrition. It is symptomatic of a much deeper malaise that cuts deeper into the innards of organizations. Attrition rates vary from sector and industry to industry.

Apart from the unavoidable ones like resignation, retirement, death or disability, the causes are found to be many and varied. They vary according to the nature of business, the level of the employees and the nature of the responsibility should

dered by them. The obvious, common and main reasons are the 'ergonomic discomfort' experienced by the employee and the 'functional incompatibility' between the corporate management and the employees. Very often an employee finds himself among colleagues and superiors he is unable to cope up with. Or he finds himself totally out of tune in his functions with the employee's functional requirements, failing to rise to the employer's expectations. Another important reason is that the employee's remuneration is not voluminous enough to bear the brunt and cushion the concussions of his family and social life.

3 Factors affecting Attrition

Some of the factors affecting attrition are:

3.1 'Money' is the culprit

Companies are wooing the best of the talents with mind-boggling salary levels and attractive designations; therefore an employee leaves an organization for another for a few thousands of rupees.

3.2 Lack of Career Mobility and Challenges

Given the choice between money and a challenging job, many employees may still prefer the latter as it allows them an opportunity to broad base their domain expertise and also provides an opportunity to work with cutting-edge technology. If the organizations do not deliver on these expectations, employee exodus cannot be contained.

3.3 Working Environment

An employee may leave an organization if the fairness of the system does not inspire his/her confidence.

3.4 High Levels of Stress and Lack of Work-Life Balance

Companies in the zeal to squeeze out every little ounce of productivity from the employees and further increase profitability. Sooner or later this makes employees stressed out and they rethink about their priorities and join an organization that promises a relaxed pace work and a breathing space.

- Brijesh Goswami is Asst. Prof. in Department of Humanities, Hindustan College of Science and Technology, Farah, Mathura, India, PH-+919917294048. E-mail: b_goswami2@yahoo.co.in
- Sushmita Jha is currently a Research Scholar in Department of Commerce and Business Administration in Central University, Allahbad, India, PH-+918938811189. E-mail: sushmita60485@gmail.com

3.5 Lack of Confidence in Supervision

It may sound cliché but the fact is that many employees leave an organization because of the immediate superior. If the supervisor lacks competence, empathy and trust in employees, they will fail to command the respect of the subordinates. (1).

3.6 Lack of Employee-job Fit

Employee's innate talent & aptitude are given a short shrift. There is a widespread notion that the employee's natural flair *per se* is not as important as new skills and knowledge acquired on the job: that with the learning attitude and training employees can do wonders in any job, but its not true and it creates immense frustration in employees, as employees demonstrate a good deal of commitment, job satisfaction, self motivation and productivity when they are assigned a job that is in tune with their natural talents.

3.7 People don't get integrated

Most organizations have an orientation program which is more of data-dump or focused on compliance trainings being completed. The focus should be more on enabling employees to form networks within themselves.

3.8 Goal setting unscientific & performance goals are unclear

Most organizations impose a normal curve fitment, but do not train managers to set realistic goals or goals that tie up with organizational or functional goals and rarely are performance goals thought through and employees told as to which resources to approach for help.

3.9 Lack of Role Clarity

Nothing can be more frustrating or discouraging for an employee than the lack of a clear understanding of what is expected of him on the job. In a performance driven workplace a lack of clarity regarding job duties and expectations can cause fear and anxiety among employees who are unclear of what is expected of them. Even worse outright anger can occur when a team member receives a negative performance evaluation based on expectations and job duties that he or she was unaware of or unclear about.

3.10 Lack of Proper Feedback

Most team members enjoy frequent feedback about how they performing. Shortening the feedback loop will help to keep performance level high and will reinforce positive behavior. But now-a-days firms fail to realize the importance of feedback mechanism, which in turn creates dissatisfaction among workforce.

4 ATTRITION SCENARIO IN INDIA

Almost all the sectors in India are facing attrition, but the reasons and effects are unique to each sector. The attrition rate in different sectors are depicted in Figure 2, While the IT Enabled services has the highest attrition rate, other sectors that are currently facing significantly high attrition are Telecom, Pharmaceuticals, Retail etc. As a result, every organization is

now striving to put into practice several measures to counter attrition, right from identifying the root cause to making the right hires. Though there is no specific data available on the gender differences in attrition levels, it has been found that women employees are less likely to switch job, merely for a modest (10-15%) hike in salaries. If given the 'right environment', which implies work-life flexibility, women tend to stay longer than men in job.

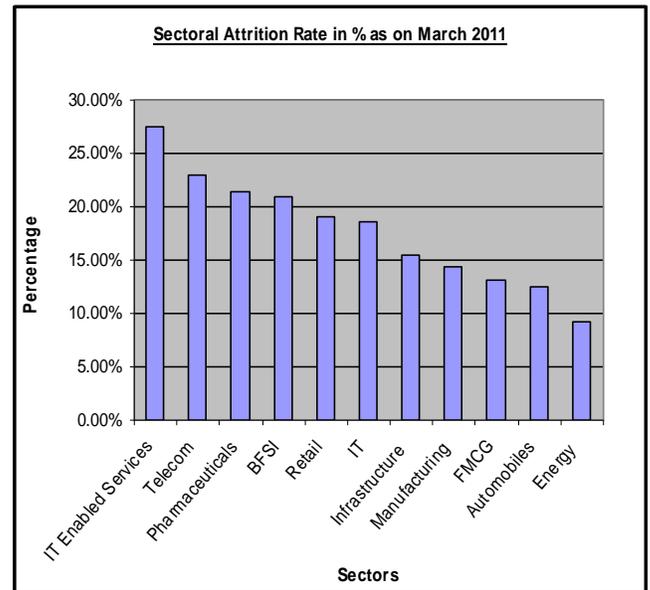


Figure 2: Attrition rate across various sectors

Source: survey conducted by HR Consultancy Firm Aon Hewitt.

5 IMPACT OF ATTRITION

The attrition of employees has a lasting impact on the firm's bottom line. When an employee quits, the need for replacement arises, the organization incurs some tangible costs. The impact ranges from costs of replacement to culture management issue with the new recruit.

5.1 Direct impact

A high attrition indicates the failure on the company's ability to set effective HR priorities. Clients and business get affected and the company's internal strengths and weaknesses get highlighted. New hires need to be constantly added, further costs in training them, getting them aligned to the company culture, etc.,—all a challenge

5.2 Indirect impact

Typically, high attrition also leads to a chronic or systemic cycle—attrition brings decreased productivity, people leave causing others to work harder and this contributes to more attrition. All this has a significant impact on the company's strength in managing their business in a competitive environment

6 POSITIVE IMPACT OF ATTRITION

Attrition is not bad always if it happens in a controlled manner. Some attrition is always desirable and necessary for organizational growth and development. The only concern is how organizations differentiate “good attrition” from “bad attrition”. The term “healthy attrition” or “good attrition” signifies the importance of less productive employees voluntarily leaving the organization. This means if the ones who have left fall in the category of low performers, the attrition is considered being healthy. Attrition rates are considered to be beneficial in some ways.

- If all employees stay in the same organization for a very long time, most of them will be at the top of their pay scale which will result in excessive manpower costs.
- When certain employees leave, whose continuation of service would have negatively impacted productivity and profitability of the company, the company is benefited.
- New employees bring new ideas, approaches, abilities & attitudes which can keep the organization from becoming stagnant.
- There are also some people in the organization who have a negative and demoralizing influence on the work culture and team spirit. This, in the long-term, is detrimental to organizational health.
- Desirable attrition also includes termination of employees with whom the organization does not want to continue a relationship. It benefits the organization in the following ways:
 - It removes bottleneck in the progress of the company
 - It creates space for the entry of new talents
 - It assists in evolving high performance teams
- There are people who are not able to balance their performance as per expectations, lack potential for future or need disciplinary action. Furthermore, as the rewards are limited, business pressures do not allow the management to over-reward the performers, but when undesirable employees leave the company, the good employees can be given the share that they deserve.

Some companies believe attrition in any form is bad for an organization for it means that a wrong choice was made at the beginning while recruiting. Even good attrition indicates loss as recruitment is a time consuming and costly affair. The only positive point is that the realization has initiated action that will lead to cutting loss

7 ANALYZING THE IMPACT

Productivity and profitability are both impacted, either negatively and positively, according to the type of attrition. Even good attrition is indicative of loss as recruitment is a time consuming and costly affair. “It is tantamount to investment that has gone astray. Having said that, good attrition minimizes the adverse impact on business while bad attrition accentuates the loss,” stated Nair. The cost of hiring is sometimes not less

than two to three times the salary of the employee.

The impact on work progress is tremendous, particularly if a project is underway and one of the key people leaves. “It leads to dip in entire organizational efficiency, and a lot depends on how it is able to cover the setback,” pointed out Rao.

Organizations should execute top of the line retention policies in the right earnest and consistency. They should be more employee-centered and look for further ways to “bond” employees to their companies. “Company performance is optimally aligned to the skills its employees possess. High attrition implies that certain necessary skills are vulnerable or are not present due to employees being lost. This results in lower than optimal levels of business performance. If the skills are constantly not available, the situation gets compounded into a crisis with key projects, revenues, etc., getting affected. Business is then reduced to just managing crisis,” added Bhardwaj. For example, a 2 percent attrition difference can make a significant difference in the ultimate business impact.

8 RETENTION CHALLENGES

Employee retention refers to policies and practices companies use to prevent valuable employees from leaving their jobs. How to retain valuable employees is one of the biggest problem that plague companies in the competitive marketplace. Not too long ago, companies accepted the “revolving door policy” as part of doing business and were quick to fill a vacant job with another eager candidate. Nowadays, businesses often find that they spend considerable time, effort, and money to train an employee only to have them develop into a valuable commodity and leave the company for greener pastures. In order to create a successful company, employers should consider as many options as possible when it comes to retaining employees, while at the same time securing their trust and loyalty so they have less of a desire to leave in the future.

Employees need to be retained because good, faithful, trained and hard working employees are required to run business. They have acquired good product knowledge over the long run and a trained employee can handle customers better and also solve problems of peers who are new to the organization. When an employee leaves he takes away with him all company information such as ongoing projects, etc. Goodwill of the company gets hampered due to more employee turnover rate and the competitors start poking their nose to recruit best talents from them. Efficiency of work is hampered to a large extent. Let me give you an example – If an employee leaves in the middle of an ongoing project its very difficult to fill that Vacuum and a new employee can never replace an old and talented employee so this leads to delayed completion of projects and less work satisfaction among other team members

9 RETENTION STRATEGIES....

When we are aware of the facts leading to attrition of an employee we can take few measures to retain them.

They are –

1. Provide good working conditions and introduce positive

work culture so that an employee is at ease and not stresses while working.

2. Appreciate all the good work and mentor him where he his wrong. He will feel at ease and will be delighted that the company cares for him.

3. The management should check that there is no miscommunication between the top management and employees because a wrong message is really dangerous.

4. Flexible working hours for those who really need to fulfill other important personal obligations.

5. Work life balance is very much necessary because if an employee has personal conflict will not work and finally leave.

6. Bonus should be given occasions and pay should be raised if an employee does good work.

7. Provide medical benefits.

8. If possible provide him with an insurance policy so that he feels that company cares for him.

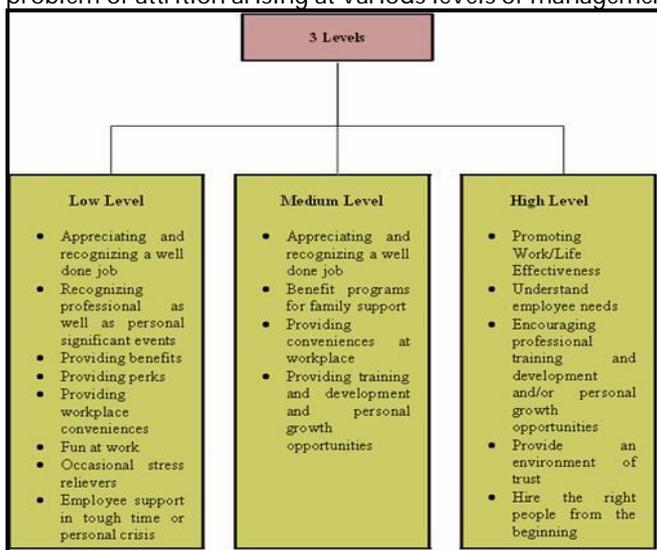
9. Recruit right person for the right job because a mismatch will hamper the work and cost a lot to the company.

10. Train the employees from time to time so that their skills enhance and they serve the organization better.

11. Have faith in your employees and include them in important decision making so that they feel they are part and parcel of the company.

Each and every company has to understand the needs of its employees and work accordingly. One approach cannot fit all, so each and every company should devise its own methods of retaining employees apart from what is always practiced.

Figure 3, throws light on the retention strategies that may be adopted to retain employees in all the three levels of management i.e. Top, Middle and Lower levels, since every level requires a different treatment because of the difference in designation, status in the organization and varied pay structures. Adopting the right kind of strategy can help to combat the problem of attrition arising at various levels of management



Source:<http://retention.naukarhub//retention-strategies.html> (16)

Figure 3: Retention Strategies for different levels of Management

The figure 3 above clearly explains that how money and understanding can create a congenial environment specially for the lower managerial level, whereas for the middle level managers appreciation of their work and various benefit programs for their families and training programs act as major attrition blocker. The high level management retention strategies includes creation of trustworthy environment and freedom of decision making, thus by understanding the needs of various levels of management an effective retention strategy can be formulated

10 CHALLENGES AHEAD.....!!!!

Job Attrition is an ongoing phenomenon and has to be dealt with psychologically. Giving a hefty pay need not satisfy an employee totally. It is the responsibility of the HR Department of the company to look at factors beyond immediate pay and find out the motivational factors giving job satisfaction to the employees. The companies have to try to create a congenial work environment and offer a promising scope for career growth for the promising performers. The HR Department should be made the part of the corporate strategy making team and a comprehensive HR Policy should be framed by experts, if necessary taking the advice of an outside consultant. Organizations should set a standard while deciding employee compensation levels and improve the reward structure year after year.

Involvement of employees in the decision-making process is the path to the feeling of emotionally engaged outcomes such as a feeling of being an important and integral part of the company. Psychometric tests should be included in the interview process to assess the behavior of the candidates. An internal program to receive employees' feedback through a feedback form at the time of bi-annual appraisal should be implemented. It offers to better rate the management performance and strategies can be carried out accordingly. Some organizations are employing handicapped people. This is profitable for organizations since they are known to be more loyal and accept a lesser pay than their physically able colleagues.

11 SOME RECOMMENDATIONS FOR EFFECTIVE RETENTION

11.1 Employee Stock Option Plans/Schemes (ESOP/ESOS)

The employee stock option plan is a good management tool for retention of human talent and guarding against poaching of staff of a running organization by a rival company. ESOP in India is regulated by Securities Exchange Board of India (SEBI). According to guidelines issued by SEBI "employee stock option" means the option given to the whole-time Directors, Officers or employees of a company which gives such Directors, Officers or employees, the benefit or right to purchase or

subscribe at a future date, the securities offered by the company at a predetermined price.

11.2 Stock Appreciation Rights (SARS)

Under an SARS, a certain number of shares are allotted to the employees. After the vesting period, the employee is free to exercise his option, and if the value of the shares has appreciated he is given shares or cash worth this appreciated value. If he is allotted shares equivalent to his appreciated value, he can sell them after a lock-in-period. The advantage of SARS is that it is a cashless transaction for the employees.

11.3 Sweat Equity

When a company is newly formed or starts a new line of business, the company engages the best executives and employees available, who bring in their IPR (Intellectual Property Rights) and know-how, skill and expertise with them, which make a value addition for the company. Certain key professionals would like to invest in the company's capital and would like to risk their own contribution to the capital of the company along with their own IPR, know-how, skill and expertise. Such employees would like to be a strategic part of the promoter group and would like to make value addition to their capital invested in the company. Such an employee is awarded with Sweat Equity as an incentive to join the company.

Sweat equity shares to mean equity shares issued by the company to employees or directors at a discount or for consideration other than cash for providing the know-how or making available rights in the nature of intellectual property rights or value additions, by whatever name called. It is, therefore, necessary for the issue of sweat equity shares that the concerned employee either provides the know-how, intellectual property rights or other value additions to the company.

11.4 'Golden Hello' and 'Golden Handcuffs'

The golden hello, as the name suggests, is an incentive offered to a select set of new recruits, mostly middle and senior-level executives poached from rivals who are well known in industry or are skilled in a niche area. The bonus could be in the form of ESOPs, cash incentives, non-cash incentives or a combination of all three. Golden Handcuffs: Any terms or restrictions that make it prohibitively expensive for the executive to leave employment, usually because he or she would be forfeiting significant stock options, restricted stock or other type of bonus or equity.

11.5 Retention Bonus

Retention bonus is an incentive paid to an employee to retain them through a critical business cycle. Retention bonuses have proven to be a useful tool in persuading employees to stay. A retention bonus plan is not a panacea. According to a survey, non-management employees generally receive about 10 percent of their annual salaries in bonuses, while management and top-level supervisors earn an additional 50 percent of their annual salaries. While bonuses based on salary percentages are the generally used, some companies choose to pay a flat figure. In some companies, bonuses range from 25 percent

to 50 percent of annual salary, depending on position, tenure and other factors. Employees are chosen for retention bonuses based on their contributions to management and the generation of revenue. Retention bonuses are generally vary from position to position and are paid in one lump sum at the time of termination. A retention period can run somewhere between six months to three years. It can also run for a particular project. As long as the project gets completed, the employees who have worked hard on it are entitled to receive the retention bonus. For example, the implementation of a system may take 18 months, so a retention bonus will be offered after 20 months.

4 CONCLUSION

Organizations planning for the future should be giving close attention to why attrition is occurring in the present. To ignore why people are leaving the organization is to ignore the organization's greatest asset – its people. People are needed to accomplish the task, but people are more than just tasks they perform. They are dreams, hopes, ambitions, creativity and innovation. To recognize and cultivate these valuable assets is one of the surest ways to build an organization that leads rather than follows in domestic and global markets.

Thus, Organizations should create an environment that fosters ample growth opportunities, appreciation for the work accomplished and a friendly cooperative atmosphere that makes an employee feel connected in every respect to the organization. Retention plans are an inexpensive way of enhancing workplace productivity and engaging employees emotionally. Proficient employees keep the quality up and business operations run smoothly along with the cost saving in the longer run paper.

ACKNOWLEDGMENT

The authors wish to thank Mr. Manish Sharma, Asst. Prof. Mechanical Engineering Department of Hindustan College of Science and Technology for his extended support in editing the article.

REFERENCES

- [1] Revisiting Empowerment as a Retention Strategy: A Snapshot – Radha Mohan Chebolu, HRM Review, March 2008 Issue, Pg. no. 53-58, ICFAI University Press, Hyderabad.
- [2] Employee Attrition: An Overview of the causes and remedies – K. Mallikarjunan, HRM Review, March 2008 Issue, Pg. no. 39-42, ICFAI University Press, Hyderabad.
- [3] Attrition is Indian firms' new worry – Vivek Sinha, Hindustan Times, (March 10, 2011) Lucknow Edition.
- [4] Attrition Control – Kiran Kumar, HRM Review, March 2008 Issue, Pg no. 43-46, ICFAI University Press, Hyderabad.
- [5] Employee Attrition and Retention: Exploring the Dimensions in the urban centric BPO Industry – Santoshi Sengupta, Jaypee Institute of Information & Technology, Noida.
- [6] Combating Attrition: An emerging Challenge in the IT Sector – Dr. Rashmi Joshi, HRM Review, March 2008, Pg. no. 47-51, ICFAI University Press, Hyderabad.
- [7] A Study on Employee Attrition and Retention in BPO Sector – Jnana Ranjan Pati (2009), (Project report of IILM Institute of Higher education), Gurgaon.

- [8] Employee Retention: An art of keeping the people who keep you in business – Priyanka Rawal, HRM Review, April 2011 Issue, Pg no. 10-17, ICFAI University Press, Hyderabad.
- [9] Attrition: From Corporate Nightmare to Competitive Advantage – Indranil Banerjee, HRM Review, March 2008 Issue, Pg. no. 68-72, ICFAI University Press, Hyderabad
- [10] 'Firms tempt talents with Golden Hellos'- Writankar Mukherjee, Economic times. (May 23, 2010).
- [11] Attrition's a Blessing for ITeS Companies – Devina Sengupta, Economic times. (August 9, 2011)
- [12] Attrition: A Global Problem- Sabitha Niketh, HRM Review, March 2008 Issue, Pg. no. 64-67, ICFAI University Press, Hyderabad
- [13] Raon V.S.P. - Human Resource Management, Second Edition, Excel Books, NewDelhi.
- [14] Sharma A.M., Understanding Wage & Compensation System, Tenth Revised Edition (2010), Himalaya Publishing House, Mumbai.
- [15] Strategies to retain Human Capital in Business Process Outsourcing Industry, (Working Paper 2) - R. Raman (February 2006), Kohinoor Business School, Khandala.
- [16] <http://retention.naukarhub//retention-strategies>

Hybrid Evolutionary Algorithm Based on PSO to Reduce Non linear effect for 802.11a High Speed Network

Er. Sahibinder Singh ¹, Er. Devesh Mahor ²

¹Deptt.Of Electronics and Comm., S.R.I.E.T, URJANI

² Deptt. of Electronics and Comm., M.M.U, MULLANA

[¹er.sahibinder@gmail.com](mailto:er.sahibinder@gmail.com)

[²mahor.devesh@gmail.com](mailto:mahor.devesh@gmail.com)

-----ABSTRACT-----

A high speed network 802.11a (OFDM) has emerged as very popular wireless transmission technique in which digital data bits are transmitted at a high speed in a radio environment. But non linear effect (PAPR) is challenge of high speed based communication that has recently drawn much attention. Non linear effect in 802.11a causes nonlinear distortions after amplified by power amplifier. Many methods proposed to reduce PAPR. In this article, we introduce a low-complexity partial transmit sequence (PTS) based on particle swarm optimization (PSO) algorithm is presented for the low computation complexity and the reduction of the peak-to-average power ratio (PAPR) of an 802.11a wireless standard. In this paper, we work around potentially computational intractability; the proposed PSO scheme exploits heuristics to search the optimal combination of Cumulative Distribution Function (CDF) with low complexity. Simulation results show that the new technique can effectively reduce the computation complexity and PAPR reduction.

Keywords: OFDM, PAPR, PSO, CDF.

INTRODUCTION:

Over the last two decades, wireless communications have been excitedly accepted by the world's population at large, to become an essential tool in our day-to-day lives. In an OFDM scheme, a large number of orthogonal, overlapping, narrow band sub-channels or subcarriers, transmitted in parallel, divide the available transmission bandwidth. The separation of the subcarriers is theoretically minimal such that there is a very compact spectral utilization. The attraction of OFDM is mainly due to how the system handles the multipath interference at the receiver. Multipath generates two effects: frequency selective fading and intersymbol interference (ISI). The "flatness" perceived by a narrow-band channel overcomes the former, and modulating at a very low symbol rate, which makes the symbols much longer than the channel impulse response,

diminishes the latter. Using powerful error correcting codes together with time and frequency interleaving yields even more robustness against frequency selective fading and the insertion of an extra guard interval between consecutive OFDM symbols can reduce the effects of ISI even more. Thus, an equalizer in the receiver is not necessary.

There are two main drawbacks with OFDM, the large dynamic range of the signal (also referred as peak-to average [PAR] ratio) and its sensitivity to frequency errors. These in turn are the main research topics of OFDM in many research centres around the world.

PAPR Problem in OFDM:

One reason why the nonlinearity of PA should be considered seriously is that the large peak power of the OFDM signal sometimes makes the

PA inefficient. When adding up subcarriers with the same phases, the peak power is N times than the average power of the signal on each subcarrier. This results in a high Peak-to-Average Power Ratio (PAPR). Such high PAPR problem associated with multicarrier signals is one of the principal drawbacks of OFDM. A high PAPR makes the PA work with large IBOs, resulting in inefficient use of the amplifier. High PAPR also increases the complexity of the ADC and DAC [21]. The PAPR is defined as:

$$PAPR = \frac{\max_{0 \leq t \leq T} |x(t)|^2}{E[|x(t)|^2]} \text{ -----(2.1)}$$

Where $\max_{0 \leq t \leq T} |x(t)|^2$ is the maximum power of the signal and $E[|x(t)|^2]$ is the average power. Another factor used is the Crest Factor (CF) which is defined as the square root of PAPR:

$$CF = \frac{\max_{0 \leq t \leq T} |x(t)|}{E[|x(t)|]} \text{ -----(2.2)}$$

From Eq. (2.1), it can be seen that the high PAPR can be reduced either by reducing the maximum signal power or by increasing the average power. In reality, reducing the maximum signal power is used in most cases because increasing the average power causes more interference. Although there are many techniques for reducing high PAPR, all these approaches have some corresponding disadvantages, such as signal distortion and complexity of the implementation. These approaches also cannot guarantee that the signal after processing can avoid PA nonlinear distortion. Here we use PSO technique for reduction of PAPR.

Introduction to PSO:

Particle Swarm Optimization was firstly introduced by Dr. Russell C. Eberhart and Dr. James Kennedy in 1995. As described by Eberhart and Kennedy [22], PSO algorithm is a population based search

algorithm based on the simulation of the social behaviour of birds within a flock. The initial intent of the particle swarm concept was to graphically simulate the graceful and unpredictable choreography of a bird flock, with the aim of discovering patterns that govern the ability of birds to fly synchronously, and to suddenly change direction with a regrouping in an optimal formation. From this initial objective, the concept evolved into a simple and efficient optimization algorithm.

PSO Algorithm.

In PSO, individuals are referred to as particles, which are “flown” through hyper dimensional search space [24]. Change in the position of each particle within the search space is based on the social psychological tendency of particle to emulate the success of other particle. The change to a particle’s position within the swarm is therefore influenced by the past experience, or by the knowledge of its neighbours. The search behaviour of a particle is thus affected by that of other particles within the swarm (PSO is therefore a kind of symbiotic cooperative algorithm). Particle Swarm optimization technique has mainly two primary operators:

- Velocity update
- Position update

During each generation each particle is accelerated toward the particle’s previous best position (pbest) and the global best (gbest) position and new velocity value for each particle is calculated based on:

- ✓ Its current velocity.
- ✓ The distance from its previous best position.
- ✓ The distance from the global best position.

The new velocity value is then used to calculate the next new position of the particle in the search space. In PSO, initially each potential solution is assigned a randomized velocity and is “flown” through the problem space. Each particle adjusts its flying according to its own flying experience and its companion flying experience.

$$v_i^{t+1} = w \cdot v_i^t + c_1 \cdot [pbest_i^t -] + c_2 \cdot [gbest -] \quad \text{-----}(3.1)$$

$$X_i^{t+1} = X_i^t + v_i^t \quad \text{-----}(3.2)$$

Where;

v_i^t is velocity of i^{th} particle at iteration t ,
 w is weight inertia.

c_1, c_2 is Acceleration Constants.

r_1, r_2 is random number between 0 and 1.

x_i^t is current position of i^{th} particle at iteration

$t, pbest_i$ is personal best of i^{th} particle.

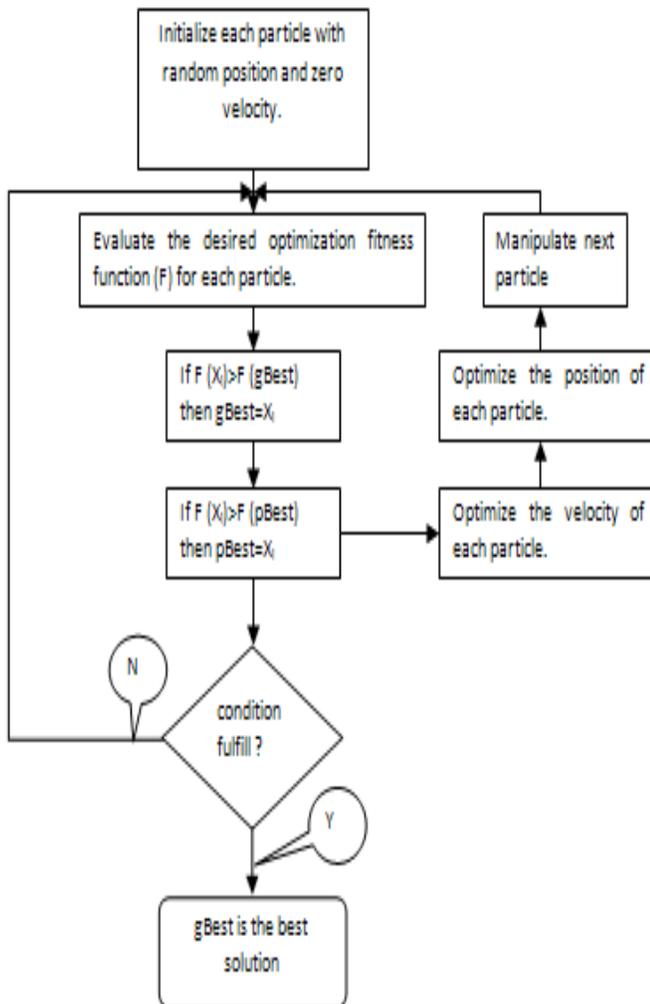


FIG:3.1 Flow diagram of PSO

Table 3.1: PSO Procedure

```

Repeat
for i = 1 to number of particles do
  if G(Xi) > G(pbesti) then G() evaluates goodness
    for d = 1 to dimensions do
      pbesti = Xi // pbesti is the best state found so far
    end for
  end if
  gbest = i // arbitrary
  for j = indexes of neighbours do
    if G(pbestj) > G(gbest) then
      gbest = j // gbest is the index of the best performer in the neighbourhood
    end if
  end for
  for d = 1 to number of dimensions do
    Vit = f(Xi(t-1), Vi(t-1), pbestj, gbest) //Update velocity
    Vi □ (-Vmax + Vmax)
    Xit = f(Vit, Xi(t-1)) //Update position
  end for
end for
until stopping criteria
end procedure
  
```

RESULTS AND SIMULATION PERFORMANCE: CUMULATIVE DISTRIBUTION FUNCTION

The Cumulative Distribution Function (CDF) is one of the most regularly used parameters, which is used to measure the efficiency of any PAPER technique. Normally, the Complementary CDF (CCDF) is used instead of CDF, which helps us to measure the probability that the PAPER of a certain data block exceeds the given threshold.

By implementing the Central Limit Theorem for a multi – carrier signal with a large number of sub-carriers, the real and imaginary part of the time – domain signals have a mean of zero and a variance of 0.5 and follow a Gaussian distribution. So Rayleigh distribution is followed for the amplitude of the multi – carrier signal, where as a central chi-square distribution with two degrees of freedom is followed for the power distribution of the system

PARAMETERS USED FOR SIMULATION:

M=4,QPSK signal constellation

No. of data points=128

Block size=08 ,size of each OFDM Block

CP_len=ceil(0.1*block size),lenth of cyclic prefix

No. of IFFT points =block size,128 points for FFT/IFFT

No. of fft points =block size.

PAPR is a big problem as far as OFDM is concerned so a method has been implemented using PSO. Initial data has been taken while doing this experiment as follow QPSK signal constellation 4, data points 128, size of each OFDM block 8, and 128 points for the FFT/IFFT. Different number of Particles (NOP) and number of iterations (NOI) have been chosen for PSO algorithm.

Figure 4.1 Shows the discrete input data which is transmitted using OFDM.

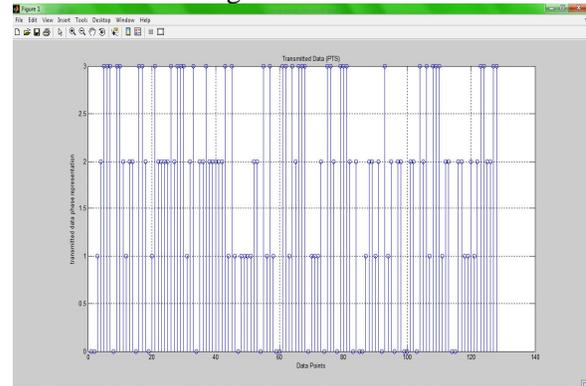


Figure 4.2 Shows the scattered plot of the transmitted QPSK modulated signal.

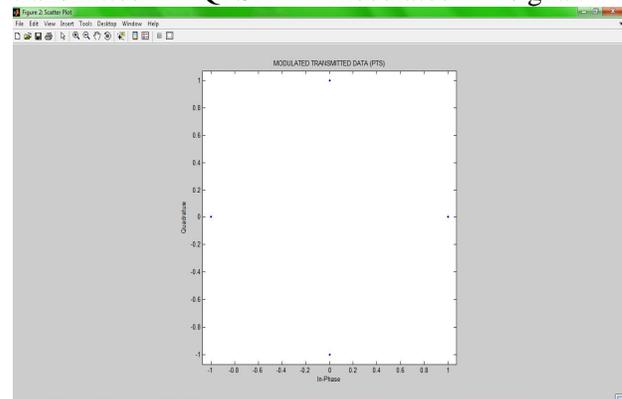


Figure 4.3 Shows the OFDM signal for the simple case which is the actual signal which transmit through the channel as only continuous signals can be transmitted.

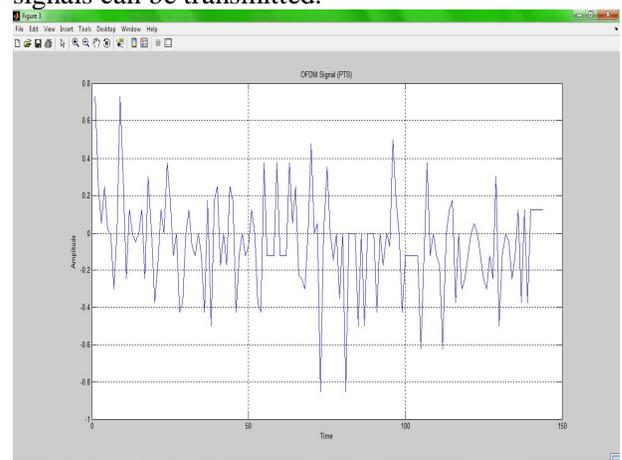


Figure 4.4 Shows the received signal for the simple case.

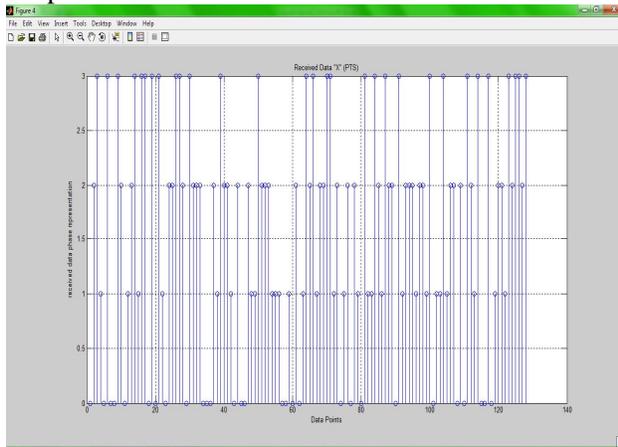


Figure 4.5 Shows the received signal for the PSO.

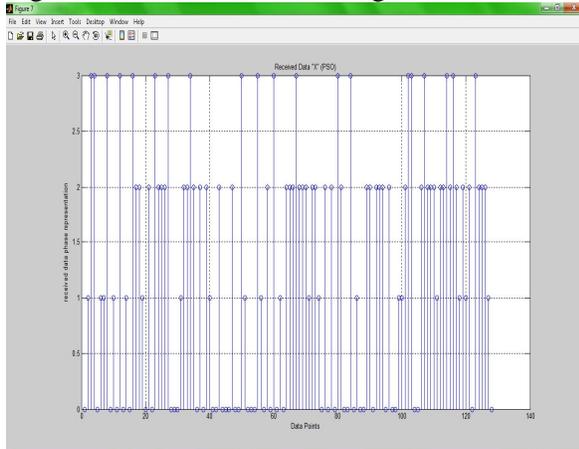


Figure 4.6 show the graph between PAPR and CDF for the number of particles and number of iteration is equal to one. Here comparison of three techniques (PTS, Clipping and filtering, proposed PSO technique) has been made on the basis of PAPR. Here for the numbers of particles and number of iteration equal to one PAPR by using PTS is 6.2, for clipping and filtering 9.1 and for PSO 10.1 for CDF equals to 0.1. So no reduction in PAPR found using PSO.

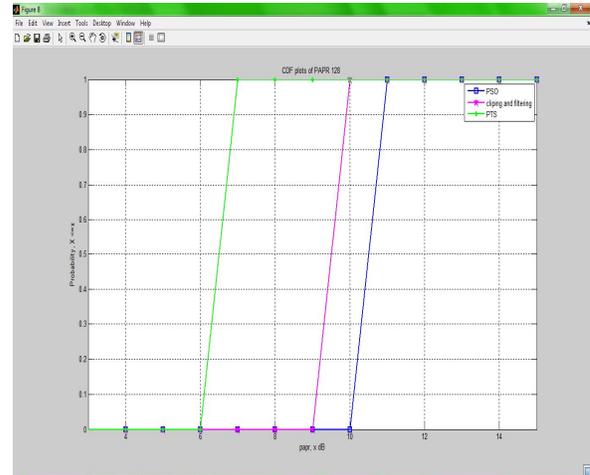
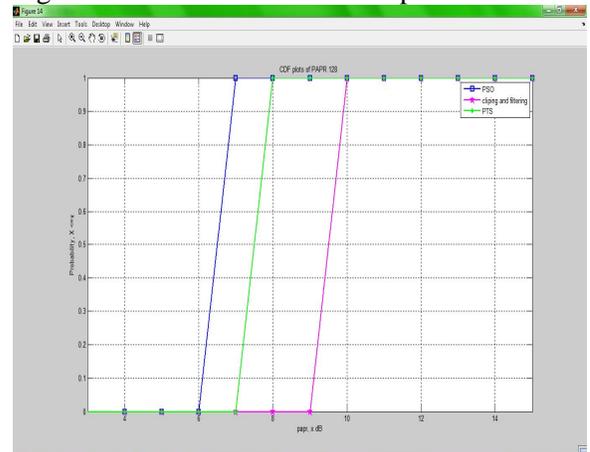


Figure 4.7 shows the graph between PAPR and CDF for the number of particles and number of iteration is equal to two. Anyone can see that for the numbers of particles and number of iteration equal to one PAPR by using PSO is 6.1, for clipping and filtering 9.1 and for PTS 7.1 for CDF equals to 0.1. So reduction in PAPR found using PSO in OFDM.

Figure 4.7 CDF V/S PAPR for $N_p=N_i=2$



OFDM using clipping not always gave good results as far as PAPR is concern for number of particles and numbers of iteration equals to two clipping gave good results but not always. As PSO is a simple PAPR reduction technique which may be easily implemented in hardware. PTS also a good technique to reduce PAPR but need modification on transmitter as well as receiver side but PSO need on transmitter only.

CONCLUSION:

With the rising demand for efficient frequency spectrum utilization, OFDM proves invaluable to next generation communication systems. To conclude, several techniques to reduce PAPR have been proposed. In recent years, such as clipping, coding and scrambling techniques. As a scrambling technique, partial transmit sequence (PTS) is known to achieve high PAPR reduction with a small amount of redundancy. However, selecting the optimal parameters for PTS is very complex, especially when a large number of sub-blocks are used. Clipping is a good technique for the reduction of PAPR but not always because some part of signal clipped that threshold should have to be chosen properly which is generally not chosen properly. But in case of PAPR reduction using PSO there is no such limit. PAPR reduction using PSO gave good result as compared of simple clipping technique and should have to be used in OFDM.

References:

1. Y Wu, W.Y Zou. "Orthogonal frequency division multiplexing: a multicarrier modulation scheme", *IEEE Transactions on Consumer Electronics*, vol.41 (3), PP. 392-399, 1995.
2. J Armstrong. "Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering", *Electronics Letters*, vol. 38(5), pp. 246-247, 2002,.
3. H.S Chen, H.Y Liang. "Combined selective mapping and binary cyclic codes for PAPR reduction in OFDM systems", *IEEE Transactions on Wireless Communications*, vol. 6(10), pp. 3524-3528, 2007.
4. R.W Bauml, R.F.H Fischer, J.B Huber. "Reducing the peak to average power ratio of multi carrier modulation by selected mapping", *Electronics Letters*, vol. 32(22), pp. 2056-2057,1997.
5. L Yang, R.S Chen, Y.M Siu. "PAPR reduction of an OFDM signal by use of PTS with low computational complexity", *IEEE Transactions on Broadcasting*, vol. 52(1), pp. 83-86, 2006.
6. D.W Lim, S.J Heo, J.S No. "A new PTS OFDM scheme with low complexity for PAPR reduction", *IEEE Transactions on Broadcasting*, vol. 52(1), pp. 77-82, 2006.
7. A.D SJayalath, C Tellambura. "Adaptive PTS approach for reduction of peak-to-average power ratio of OFDM signal", *Electronics Letters*, vol. 36(14), pp. 1226-1228, 2000.
8. H.H Seung, H.L Jae. "PAPR reduction of OFDM signals using a reduced complexity PTS technique", *IEEE Signal Processing Letters*, vol. 11(11), pp. 887-890, 2004.
9. L.J.Jr Cimini, N.R Sollenberger. "Peak-to-average power ratio reduction of an OFDM signal using partial transmit sequences", *IEEE Communications Letters*, vol. 4(3), 2010.
10. R. Van Nee and R. Prasad, OFDM for wireless multimedia communications, Artech House, 2000.
11. D. Petrovic, W. Rave and G. Fettweis, "Effects of Phase Noise on OFDM systems With and Without PLL: Characterization and Compensation," *IEEE Trans. Communications*, vol.55, no.5, August 2007..
12. Y. Ding, Y. Liu, I. Nilkhamhang and A. Sano, "Adaptive Linearization for Power Amplifier in OFDM

Systems," SICE-ICASE, October
2006.

Author's biography:

Er. Sahibinder Singh

Pursuing M.Tech in Electronics & Comm.
Engg. From MMU, Mullana.

Working as a Lect. in Electronics & Comm.
Deptt. In Shri Ram Institute of Engg. &
Tech., Urjani

Er. Devesh Mahor

M.Tech in Electronics & Comm. Engg.
Working as Sr. Lect. in Electronics &
Comm. Deptt. in MMU, Mullana.

Enhanced VoIP Based Virtual PC Troubleshooting

U.K.D.S.N. Dayananda, W.N.K.L. Abeyrathna, S.M.C.D. Samarakoon, K.B.M.S. Jayawickrama,
M. Kalyanapala, M.N. Wijesundara

Abstract— According to the businesses and marketing strategies, customer relationship management and call center outsourcing are becoming major concerns since customer satisfaction is considered as the ultimate goal of any kind of business organization. This paper discusses how the VoIP technology can be used in an automated remote PC troubleshooting environment along with Interactive Voice Response and remote desktop assistance, for the purpose of creating ultimate software based PC troubleshooting solution capable of operating with reduced human involvement in order to provide optimum availability, reliability and cost-effectiveness for PC related business organizations. In this solution, a call processing engine pre-processes incoming calls and redirects them into the Interactive voice response engine which is enforced with a built-in PC troubleshooting knowledge base capable of guiding the caller to a solution with voice commands. If the knowledge base fails to provide a reasonable solution, the remote assistance enables real-time interaction between the caller and a human trouble-shooter. This approach minimizes the current drawbacks of traditional PC troubleshooting and increases the productivity and the efficiency.

Index Terms— PC troubleshooting, Session Initiation Protocol (SIP), Interactive Voice Response (IVR), Remote Desktop Assistance, Decision tree.

1 INTRODUCTION

With the emergence of the personal computer, human life style became more and more fascinating, simple and automotive. In 1957 the first commercial PC was introduced by IBM[1] and thereafter many PC companies were born such as HP[2], DELL[3], Apple[4] etc. While the commercial PC became a daily partner of human life, its main goal was to create the computer era and simplify the human lifestyle as much as possible. The number of personal computers in use worldwide hit one billion in June 2008, and is expected to reach 2 billion by 2014[5]. But the main issue is; is there any systematic mechanism to increase the after sales maintenance and customer care services for those PC's vendors in order to provide good customer satisfaction and an acceptable level of recovering PC related issues.

The currently accepted solutions for the above addressed problem domain is traditional PC troubleshooting mechanisms where, if a particular customer needs assistance in resolving an issue of a PC, the customer can inquire from the customer care division by phone or using online support which normally provides FAQs and forums [6],[7] dedicated for PC troubleshooting services and after that it is necessary to bring the PC to a service center of the relevant vendor organization whether the issue is in hardware or software. According to the existing business marketing strategies, the technical and maintenance services are still in their primary stages because

PC vendor companies are still expending a considerable amount of their annual budget for customer relationship management and service maintenance. A traditional call center solution is fairly unacceptable in modern state of art technologies and those are becoming bottlenecks in modern day business development strategies by wasting customer time, company's efficiency and productivity as well as reputation and degrading customer satisfaction. Here are the main problems faced by vendors related to the existing procedure;

- i. A company needs lots of staffing facilities and ground spaces for call centers or outsourcing the service to a separate company.
- ii. If the company has globally expanded their business, they need to maintain a regional or country vice service and technical centers which cost huge annual expenditure.
- iii. Initial cost of a call center is highly expensive.
- iv. It's not cost effective to expand the call center capacity.
- v. An organization is has not any standard provable way to measure the call center agents productivity.
- vi. Technical agent's skills are always not effectively used to their maximum potential.

The enhanced VoIP based virtual PC troubleshooting solution 'BackSpace', which has the ability of breaking the traditional barriers of PC troubleshooting services and used the combination of virtual call center architecture alone with interactive voice response aided automated remote PC troubleshooting procedure, so the PC vendor organizations will be able to use the state of art technologies such as VXML[8](We used VXML 2.0), VoIP[9], SIP[10] etc to improve the productivity, efficiency and customer satisfaction and achieve their business goals more confidently.

- U.K.D.S.N. Dayananda (smrdayananda@gmail.com), W.N.K.L. Abeyrathna (kasunabeyrathna@gmail.com), S.M.C.D. Samarakoon (dinethvirtusa@gmail.com), K.B.M.S. Jayawickrama (bhagya727@gmail.com) were with Sri Lanka Institute of Information Technology, Malabe, Sri Lanka.
- M. Kalyanapala (kalyanapala.m@sliit.lk), M. N. Wijesundara (malitha.w@sliit.lk) are with the Electronic and Computer Engineering Department, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka.

those operations are still handled manually.

2 METHODOLOGY

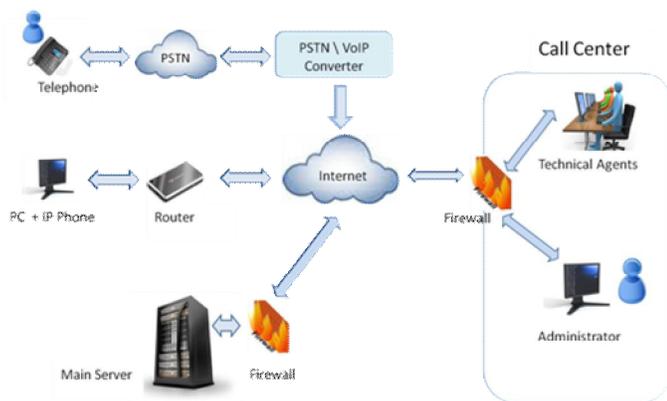


Fig. 1. System diagram

The design is mainly based on 3-tier architecture which separates the entire solution into three separate service layers. It accompanies VoIP technology to transfer voice through IP [11] and use internet and PSTN [12] as communication networks, SIP as application layer protocol and RTP [13] as transport layer protocol. It also provides virtual call center architecture so the employees can connect to the system as shown in Fig. 1. irrespective of the location and the system will be available anytime anywhere.

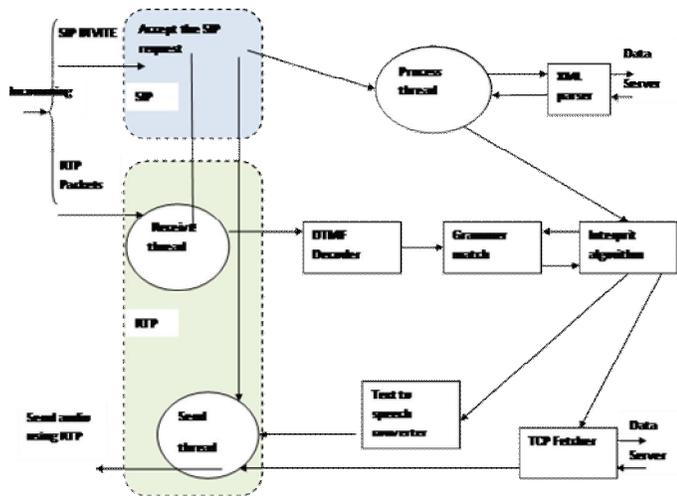


Fig. 2. Main server Process flow diagram

Fig. 2. shows the diagram of the process flow of the Main server. When the server receives an incoming request, it accepts the call and invokes the process thread and RTP thread. RTP thread receives the incoming RTP packets and invokes the DTMF [14] decoder. If any DTMF signal is detected, then it immediately sends to the grammar matching module and checks for any valid DTMF digit. The process thread is responsible of handling initial VXML page requests using the

XML parser which fetches the page to a tree structure. Then the process thread calls the interpreter algorithm and it accepts the selected page as the input. It has the ability to invoke several modules based on the VXML page contents such as invokeText to Speech synthesizer to synthesize any <prompt> or invoke another TCP request. This VXML page also can specify the grammar rules too. Send thread is used to pre-process and send the data to the caller which is coming from speech synthesizer or TCP fetcher.

The BackSpace solution is consisting of three major components;

- A. Call processing Engine and DTMF Decoding subsystem
- B. Interactive Voice Response (IVR) sub system
- C. Remote Desktop Assistance Sub System

A. Call Processing Engine and DTMF Decoding

Call processing Engine is responsible for handling all the call processing tasks and internal system monitoring activities. User friendly advanced Graphical User Interfaces (GUIs) are provided for necessary controlling and monitoring of processes.

The call processing unit provides functionalities and services to balance the overall system load as follows;

- Mapping PSTN to IP
- Call Queuing
- Call Transferring
- System monitoring and log file creation

BackSpace is designed to use some advanced features of call center architecture. It continuously accepts incoming calls from both IP phones as well as the POTS (Plain Old telephony Service) via PSTN networks. Since we are using VoIP, we converted PSTN analogue signals into SIP supported data using a FXO (Foreign Exchange Office) [15] analog voice card.

Advanced call routing and call queuing is used to handle the multiple user connections of the system simultaneously. Our implementation has the capability of handling hundred (100) simultaneous calls. When the number of calls exceeds more than hundred, the next incoming request will be queued without disconnecting. Each call is handled by creating a separate thread in order to provide unique service for each request.

The primary stage of providing a solution for a particular issue is achieved by providing automated step by step advices through a problem solving hierarchy. If the automated system is unable to solve the problem as expected by the user, due to lack of information in the PC trouble shooting knowledge base it is needed to transfer the call for human attention, as a secondary stage. To achieve this functionality we have designed and implemented a call transferring algorithm to find out an available best suitable person in order to transfer the call.

System monitoring and log file creation is added to this solution for the administration purpose. Administrators can log into the system and view the call history, agent productivity and the system records.

1) *DTMF Decoding*: DTMF decoder is primarily used for decoding the incoming analog signals of the caller and detect whether the incoming signal has any DTMF digit, if so decode it and send it to the IVR unit in order to interact and browse through the PC troubleshooting hierarchy.

When considering commercial products, all the systems use hardware based DTMF decoding solutions but in our solution we used a totally software based DTMF decoder (developed using C# 4.0) and it acts as a bridge between the caller and the IVR unit.

We have used 'Goertzel algorithm [16]' when designing the DTMF decoder because Goertzel algorithm does not require much memory and is optimal when used to compute to deal with frequencies.

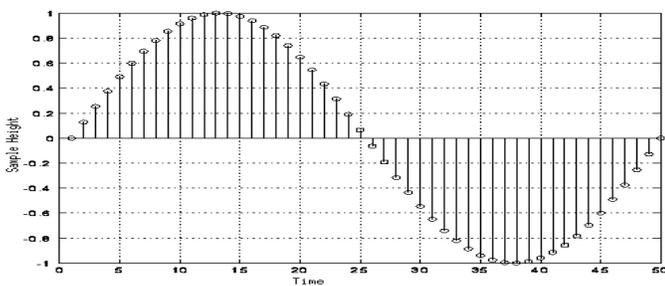


Fig. 3. Sampling of the wave signal

As Goertzel algorithm implemented is based on sampling method, we had to sample the wave signal (Fig. 3.) in to a number of discrete values (640 samples) and, using coefficient values it will detect the client pressed key.

DTMF component is responsible for performing following tasks.

- Listen for client DTMF tones
- Generate the wave signal for integer values
- Sampling the wave signal to frames
- Identifying the tone pressed by the client
- Send identified DTMF tone to IVR unit

B. *Interactive Voice Response (IVR)*

The IVR component directly communicates with the server and the DTMF decoder. There is a separate DTMF decoder specially designed for this purpose and it continually listens for DTMF data from the incoming RTP packets received by the caller's phone.

Interactive Voice Response unit is the main module which is responsible for interacting with the human actors who are connecting to the BackSpace system through PSTN telephones, IP phones or soft phones. Once the system is up, the IVR module is responsible for continuously listening on incoming client requests which come to the IVR unit as an initial caller request or as a DTMF digit. The IVR module consists of three sub components (Fig. 4.).

- SIP application server (deploys the application logic)
- Media server (IVR commands are interpreted, executed, DTMF detected and recording audio signals)
- Data Server (database for prompts and repository that stores VXML pages and other data bases)

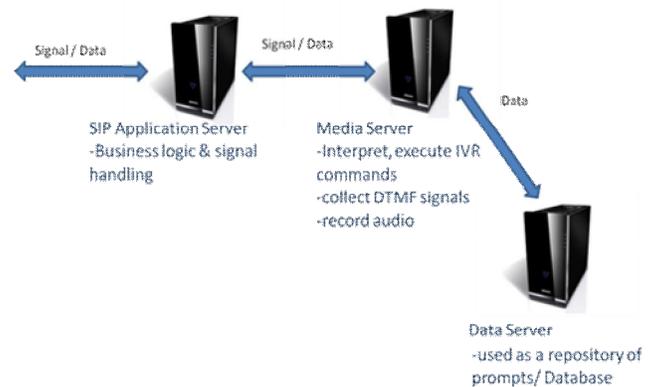


Fig. 4. Sub components of the IVR

1) *SIP Application Server*: The functionality is similar to a normal HTTP web server. But the difference in here is only the VoIP applications requests can be handled. It has the ability of taking care of multiple thread handlings, running several VoIP applications simultaneously with better memory handlings. SIP application server is responsible for accepting SIP requests from the outside. It uses port 5060 which is used for UDP [17].

2) *Media server*: Responsible for handling audio/video stream flows via RTP sessions. So the media server is the entity that interacts directly with the end-user devices in order to play prompts, record audio streams and detect DTMF digits. Here TRP flows between the media server but SIP signal does not. SIP only does the session initiation between two end points and then resides silently until the RTP session terminates.

3) *File Server*: File server is the repository of the BackSpace solution and acts as the data tier of the system. File server communicates only with the media server using TCP connection. This is a repository which is used for storing VXML pag-

es, pre-recorded audio files and SQL database. The primary objective of placing these files in a separate server is to achieve more scalability, security and reliability.

The IVR unit uses the VXML 2.0 technology to interact with the caller. The solution is capable of analysing DTMF (Dual Tone Multi Frequency) dial pad input of the caller's telephone and response with the relevant audio output. For this purpose VXML browser/parser is used. It gets the caller input, divides a particular VXML page's contents which are surrounded by <prompt> tag into its corresponding audio stream and played back to the caller. Following (Fig. 5.) is the example VXML content which fetches the user selection.

```
<grammar version="1.0" root="top" mode="dtmf">
<rule id="top">
<one-of>
<item>1</item>
<item>2</item>
<item>3</item>
<item>4</item>
</one-of>
</rule>
</grammar>
```

Fig. 5. VXML code fragment

When considering the IVR component, it addresses most drawbacks of the current call center or any other IVR systems in the market. But this is a customized version when compared to a typical IVR application; some unnecessary features which are not related to the objectives of the IVR component used in BackSpace are removed in order to achieve high quality and the simplicity of the solution.

IVR administration panel shows in Fig. 6. provides a highly user-friendly environment to customize and edit the IVR knowledge base. According to the current usage, it has the ability to customize the IVR unit into another problem domain. (As an example this solution can be transformed in to an automated call center which can be used in a travel agent company). Through this panel, a client can do all the necessary modifications.

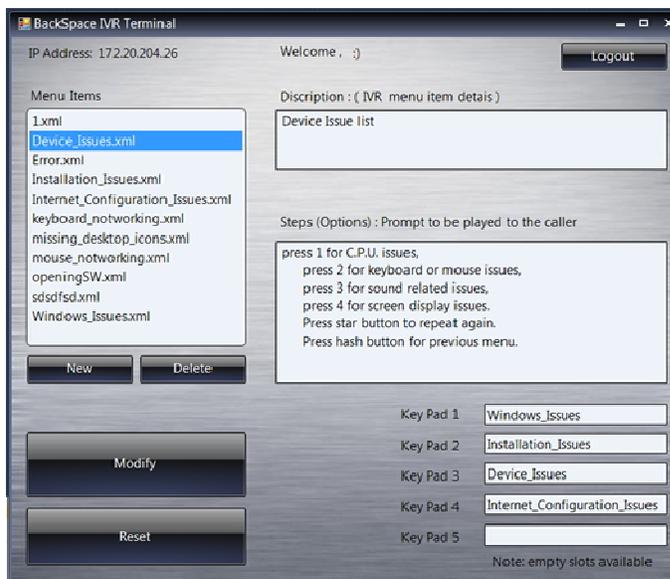


Fig. 6. IVR administration panel

C. Remote Desktop Assistance Sub System

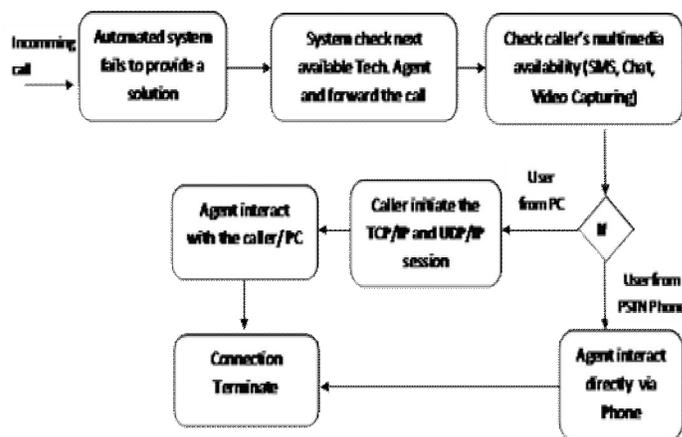


Fig. 7. Remote desktop assistance process flow

The remote desktop assistance is used to achieve the real-time communication between a caller and a technician as the secondary stage of problem solving phase of the solution. Fig. 7. shows the process flow of the remote desktop assistance. The connection initialization is done by the caller. That will be a TCP/IP connection and UDP/IP and TCP/IP connection. This design is totally implemented using .Net socket programming [18]. There are three communication methods used to interact with the caller.

- Text Communication
 - Voice Communication
 - Desktop Capturing
1. *Text Communication:* For text communication implementation we have used library called conference library [18].

Text message typed in the textbox will be sent to destination IP address using UDP/IP connection and socket programming.

2. *Voice Communication*: For chat communication implementation we have used compressed methods called Alaw [19], MuLaw [19] to compress voice data before sending over the internet. Voice data will be sent to destination IP address using UDP/IP connection and socket programming.
3. *Desktop Capturing*: Remote desktop capturing designed and developed using TCP/IP connection and socket. When sending initial Bitmap images to the other side we convert and compress the data into JPEG [20] model before sending through the network since it is very efficient and very easy to handle and process. Screen will be captured every second and if the screen is same like the previous one that will not be sent and save as bit map image for fast processing.

3 RESULTS AND DISCUSSIONS

The main goal of this research was to develop an automated software based solution for medium to large-scale PC vendor organizations to automate their traditional PC maintenance and customer care services. This solution has the ability of breaking the traditional barriers of PC related technical service providence. Furthermore, the highly customizable user-friendly environment enables the system to change its decision tree according to any business domain so it can be use anywhere as a replacement for the traditional call center. It uses the combination of virtual call center architecture along with automated remote PC troubleshooting to overcome above addressed issues, so the PC vendors will be able to use this solution to improve the productivity, efficiency and customer satisfaction and achieve their business goals.

We have successfully designed, developed and implemented BackSpace PC troubleshooting solution to handle hundred simultaneous calls and the troubleshooting decision tree contains solutions up to considerable level. When developing the solution we used SipekSdk [21] which is a C# based free and open source SDK, together with pjsip[22] (Open source SIP stack) and pjmedia[23] (Open source media stack) for VoIP based call processing and media handling. The decision tree contains only Windows XP [24] related issues and it provides automated audio instructions only in the English language. We successfully demonstrated our achievements in the final presentation of our research project.

4 CONCLUSION AND FUTURE WORK

The current solution is capable of improving its functionality and features since the whole solution is developed using a component based approach with well-defined interfaces. The

following are the key features that we are planning to implement in the near future.

1. Change the system to deploy and function in Linux environment.
2. Design and implement a VPN to connect administrators and technicians to the main server.
3. Improve the functionality and efficiency of the VXML parser.
4. Currently the PC troubleshooting knowledge base includes solutions only for Microsoft Windows XP operating system. We are planning to extend it to have troubleshooting solutions for a Linux based operating system.
5. Provide language selection facility for the callers before entering to the troubleshooting process.
6. Improve the overall security of the system.

ACKNOWLEDGMENT

'BackSpace' research team wishes to acknowledge Mr. Jayantha Amaraarachchi, Mr.Nuwan Kodagoda, Mrs. A. T. C. V. Hettiaratchy, Mr.Bhaggya Solangaarachchi and other contributors for guiding, encouraging and supporting us to make this venture a success.

REFERENCES

- [1] IBM Company, United States [Online]. Available: <http://www.ibm.com/>
- [2] HP Computers, United States [Online]. Available: <http://www.hp.com/>
- [3] DELL Computers [Online]. Available: <http://www.dell.com/>
- [4] Apple Inc, United States [Online]. Available: <http://www.apple.com/>
- [5] "HOW MANY COMPUTERS ARE IN USE IN THE WORLD TODAY AND HOW MANY PERSONAL COMPUTERS WERE SOLD IN 1977?" [ONLINE]. AVAILABLE: <HTTP://WWW.BIGSITEOFAMAZINGFACTS.COM>
- [6] HP network and internet support, HP Computers, United States. [Online]. Available: <http://www8.hp.com/us/en/support-topics/network-internet-support.html>
- [7] IBM online computer troubleshooting assistance. [Online]. Available: <http://www-947.ibm.com/support/entry/portal/Troubleshooting>
- [8] Voice Extensible Markup Language (VoiceXML) [Online]. Available: <http://www.w3.org/TR/voicexml20/>
- [9] VOIP TECHNOLOGY OVERVIEW [ONLINE]. AVAILABLE: <HTTP://MSDN.MICROSOFT.COM/ENUS/LIBRARY/AA922928.ASPX>
- [10] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler. SIP: Session Initiation Protocol. RFC 3261, Internet Engineering Task Force, June 2002.
- [11] Marina del Rey, Internet Protocol. RFC: 791. Information Sciences Institute, University of Southern California. September 1981.
- [12] Daniel Livengood, Jijun Lin and ChintanVaishnav "Public Switched Telephone Networks". Engineering Systems Division, Massachusetts Institute of Technology. May 2006 .

- [13] H. Schulzrinne, S.Casner, R.Frederick, V. Jacobson. RTP: A Transport Protocol for Real-Time Applications. RFC 3550. Internet Engineering Task Force, July 2003.
- [14] DTMF Explained [Online].
Available :<http://www.genave.com/dtmf.htm>
- [15] Understanding foreign exchange office(FXO) voice interface cards.[Online].Available :
http://www.cisco.com/en/US/products/hw/routers/ps274/products_techn_note09186a00800b53c7.shtml
- [16] THE GOERTZEL ALGORITHM. [ONLINE].AVAILABLE :
[HTTP://WWW.EETIMES.COM/DESIGN/EMBEDDED/4024443/THE-GOERTZEL-ALGORITHM](http://WWW.EETIMES.COM/DESIGN/EMBEDDED/4024443/THE-GOERTZEL-ALGORITHM)
- [17] J. Postel. User Datagram Protocol. RFC 768, Internet Engineering Task Force, August 1980.
- [18] Introduction to Sockets and Socket Programming.[Online].Available:
<http://compnetworking.about.com/od/itinformationtechnology/l/aa083100a.htm>.
- [19] **G.711 codec process [Online]Available:**
<http://www.en.voipforo.com/codec/codecs-g711-alaw.php>
- [20] CCIT, INFORMATION TECHNOLOGY-DIGITAL COMPRESSION AND CODING OF CONTINUOUS-TONE STILL IMAGES-REQUIRMENTS AND GUIDLINES. ITU, 1993.
- [21] sipeksdk.
[Online].Available:<http://code.google.com/p/sipeksdk/>
- [22] PJSIP, Open Source SIP Stack.
[Online].Available:<http://www.pjsip.org/>
- [23] pjmedia,Open Source Media Stack
[Online].Available:<http://www.pjsip.org/>
- [24] "WHAT IS WINDOWS XP", INDIANA UNIVERSITY,UNITED STATES.[ONLINE].AVAILABLE:[HTTP://KB.IU.EDU/DATA/AKMA.HTML](http://KB.IU.EDU/DATA/AKMA.HTML)

Maximizing Network Lifetime in Wireless Sensor Network Using HEF

P.Ponmegala, R.Karthik

Abstract:

Finding lifetime of sensor network is an essential part while using sensors for operation. Predicting the total lifetime of the network by Schedulability analysis and improving the network lifetime. In this paper we introduce a clustering algorithm for maximizing the lifetime of network called HEF algorithm and also adding an extra relay station to base station for effective transmission of packets. So we are avoiding the packet loss which results in saving energy for network to lead the activities to some time extend.

Key terms: Schedulability, relay station, clusters.

1. INTRODUCTION

Wireless Sensor Networks (WSNs) have a great number of nodes with sensing, computing, and wireless communication functionalities. WSNs are used in safety-critical or highly-reliable applications, two constraints are

- Real-time constraints
- Network life constraints

Two time constraints on WSN based safety critical systems time constraints. Real-time computing is the study of systems that should operate correctly under time constraints. There are two types of real-time systems:

Hard Real-time system:

It does not allow any task to miss its deadline.

Soft Real-time system:

It strives to satisfy deadline requirements statistically.

A typical HC-WSN is comprised of a base station, several cluster head nodes, and regular sensor nodes. For administrative purposes, the operation of a HC-WSN is divided into rounds in which sensor nodes are grouped into clusters. Each round consists of three phases:

- Cluster head selection
- Cluster formation
- Data communication.

The clustering process involves so many steps.

1) HEF selects cluster heads according to the energy remaining for each sensor node, and then the "setup" message (indicating cluster members, and the cluster head ID for each participated group) is sent to the cluster head of each cluster.

- 2) The cluster head of each group broadcasts the "setup" message inviting the neighbour sensor nodes to join its group.
- 3) After receiving the "setup" message at this round, the regular sensors send the "join" message to its corresponding cluster head to commit to associate with the group.
- 4) Each cluster head acknowledges the commitment, and sends TDMA schedule to its cluster members.
- 5) All sensors perform its sensing and processing and communication tasks cooperatively at this clock cycle (round). Each sensor sends its energy information to its cluster head at the end of this clock cycle.
- 6) Upon collecting cluster members' information at a given period, the cluster head sends the summarise report to the base station.

2. ALGORITHM

2.1 HEF CLUSTERING

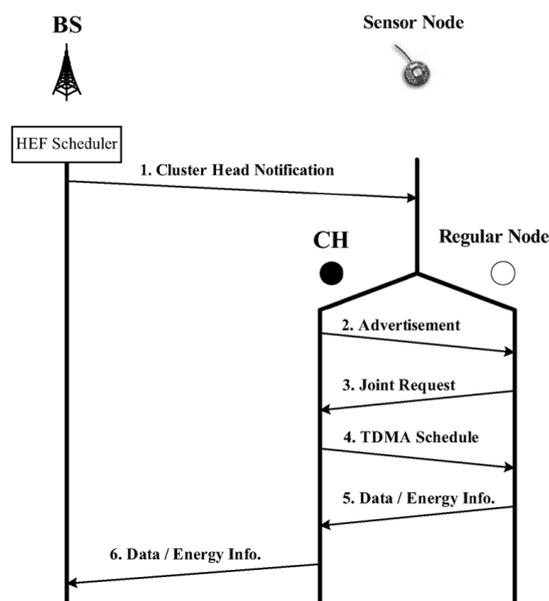


Fig 1: Information flow of the centralized HEF system.

2.2 ALGORITHM PHASES

HEF selects the set of highest ranking energy residue sensors for cluster heads at round where denotes the required cluster numbers at round. Some researchers have claimed that HEF is an efficient cluster selection algorithm that prolongs network lifetime based on simulations. However, their measurements and simulation results are stochastic processes. . Each round comprises the following three phases:

- CHS Phase,
- CFM Phase,
- DCM Phase.

2.2.1 Ideal Conditions for Optimality of HEF (ICOH):

1) All nodes must operate in a working-conserving mode. In other words, each node works as a clutter head, or a regular sensor in a round.

2) The energy consumptions of and are constant during the entire operation where, in the working-conserving mode, sensor nodes must consume energy at anytime while they operate.

2.3 HIGH-ENERGY-FIRST(HEF) HEURISTIC

Based on above observations we propose a new heuristic to solve the target coverage problem. We observe that the granularity parameter w plays an important role in getting a better approximation of optimal

solution. Hence prioritizing the sensors in terms of residual battery provides us better opportunity of using the sensors. HEF uses the three important steps in the following manner.

- Generate a cover
- Assign lifetime to a cover
- Change the priorities of the sensors

Schedulability test flowchart applies the worst-case energy consumption analysis to derive the predictability of HEF. Schedulability tests allow engineers to assess what actions (e.g. changing energy budget or lifetime, etc.) should be taken to improve the dependability and reliability of the systems. As shown in figure the Schedulability test flow chart consists of three major stages:

- Deployment Planning,
- Energy Estimation,
- Schedulability Analysis.

2.4 NETWORK LAYERS

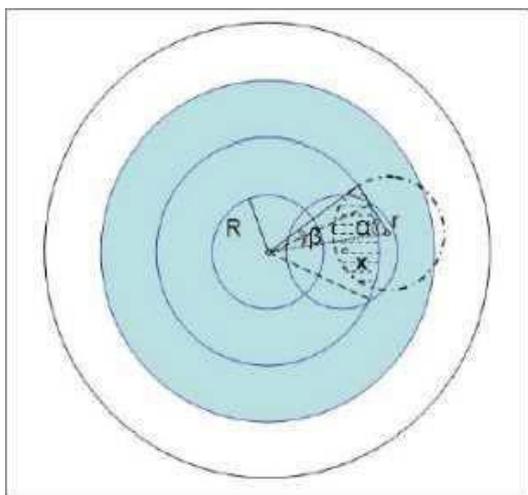


Fig 2: Sensor network divided into layers.

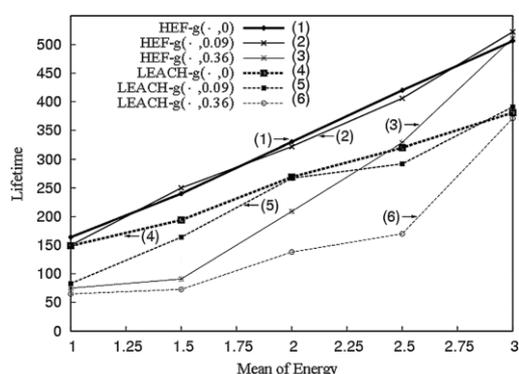
CONCLUSION

On providing a trustworthy system behaviour with a guaranteed hard network lifetime is a challenging task to safety-critical and highly-reliable WSN applications. For mission critical WSN applications, it is important to be aware of whether all sensors can meet their mandatory network lifetime requirements. In this project, we have addressed the issue of the predictability of collective timeliness for WSNs of interests. First, the High Energy First (HEF) algorithm is proven to be an optimal cluster head selection algorithm that maximizes a hard N-of-N lifetime for HC-WSNs under the ICOH condition. Then, we provide theoretical bounds on the feasibility test for the hard network lifetime for the HEF algorithm. Our experiment results show that the HEF algorithm achieves significant performance improvement over LEACH, and HEF's lifetime can be bounded. We have also developed formulas to derive the upper and lower bounds of the network lifetime quickly and easily (including both loose and sharp bounds). In particular, the feasibility test analysis performed in this paper presented a solution that would guide the system administrator to ensure that the system lifetime is predictable.

3. GRAPHICAL ANALYSIS

With the increase in initial energy, the lifetime for all schemes increases, but HEF prolongs the network lifetime as compared to LEACH when the initial energy becomes

large enough. This result is because LEACH is unable to balance the energy consumption among the sensor nodes to avoid early energy depletion of the network. When the initial energy level is low, there is no significant performance difference between HEF and LEACH. However, HEF has better performance at a small variance. The HEF algorithm performs better out of all LEACH schemes under high initial energy level. In this experiment, HEF surpasses LEACH by taking into account network lifetime when they have the same initial energy level.



3.1 TABLE I
 Simulation Parameters

Parameters	value
Number of nodes	100
Number of clusters	5
Network size	100mX100m
Base station location	(50,180)
Radio speed	1Mbps
Header size	25 bytes
Packet size	500 bytes
Radio electronics energy	50 nJ/bit
Compression ratio	0.5

4. SCOPE OF FUTURE WORK

Minimizing the packet loss by adding extra stations to receive. In addition to the base station we have use another terminal station to receive the data packets during the transmission.

5. ADVANTAGES

- Loss of packets will be reduced.
- Energy loss is avoided.
- Time consumption.

6. REFERENCES

- 1) J. A. Stankovic and K Ramamrithan, Eds., Tutorial on Hard Real- Time Systems. IEEE Computer Society Press, 1988.
- 2) B.-C. Cheng, A. Stoyenko, T. Marlowe, and S. Baruah, "LSTF: A new scheduling policy for complex real-time tasks in multiple processor systems," Automatica, vol. 33, no. 5, pp. 921–926, May 1997.
- 3) C. L. Liu and J. W. Layland, "Scheduling algorithms for multiprogramming in a hard-real time environment," Journal of the Association for Computing Machinery, vol. 20, no. 1, pp. 46–61, January 1973.
- 4) R. Cristescu and M. Vetterli, "On the optimal density for real-time data gathering of spatio-temporal processes in sensor networks," in Fourth International Conference on Information Processing in Sensor Networks (IPSN '05), April 2005.

5) W. Wang and A. Jantsch, "An algorithm for electing cluster heads based on maximum residual energy," in Proceeding of the 2006 international conference on Communications and Mobile Computing, July 03-06, 2006.

6) E. Hansen, J. Neander, M. Nolin, and M. Björkman, "Energy-efficient cluster formation for large sensor networks using a minimum separation distance," in In The Fifth Annual Mediterranean Ad Hoc Networking Workshop, Lipari, Italy, June 2006.

7) H. Chen, C.-S. Wu, Y.-S. Chu, C.-C. Cheng, and L.-K. Tsai, "Energy Residue Aware (ERA) clustering algorithm for leach-based wireless sensor networks," in Second International Conference on Systems and Networks Communications (ICSNC 2007), 2007, pp. 40–45.

8) H. Kaur and J. Baek, "A strategic deployment and cluster-header selection for wireless sensor networks," IEEE Transactions on Consumer Electronics, vol. 55, no. 4, November 2009.

AUTHOR 1



P.Ponnemegala, received B.E degree in 2010 from Anna University Chennai India. Currently pursuing M.E degree in Computer Science and Engineering in Karpagam University, India.

AUTHOR 2

R.Karthik, received B.E (CSE) degree from Anna University, Chennai, India, M.E (CSE) degree from Anna University, Thirunelveli, India. His research interests include Cloud Computing. At present, serving as Assistant Professor at Karpagam University, Coimbatore, India.

Graph Applications to Fluid Mechanics

V.Manjula

Abstract Fluid mechanics is an important subject which has been given equal weight age in Mechanical, civil, Chemical Engineering curriculum. It deals with the flow of fluids. This paper designed to explain the fundamentals of fluid mechanics in the areas of properties of fluids. And its applications to graphs A general graph approach for computational fluid dynamics is presented. Density, velocity and entropy per unit volume are used as independent variables for a single-phase, single-component flow. As a result of a combination of graph concepts with elements of Fluid mechanics, a new approach was developed principles and concepts of Fluid mechanics presented in simple and clear terms. The paper has an easy to read style and is going to benefit the readers. Important findings are a fluid is a substance which can i)attain the shape of the container as it has no definite shape. ii) Fluid is a substance which can't acquire any static equilibrium under the action of any sheer force of even a small magnitude

Index Terms— Introduction, properties of fluids, objectives, Graphical representations

1. INTRODUCTION

It is the science in which we study the behavior of fluids which are either in rest or in motion. Fluid is a substance which can flow. Technically the flow of any substance means a continuous relative motion between different particles of the substance The analysis of the behavior of fluids is based on the fundamental laws of mechanics which relate continuity of mass and energy with force and momentum together with the familiar solid mechanics properties. The study of fluids under static conditions is called Fluid statistics. Matter exists in two states; the solid and the fluid, the fluid state being commonly divided into the liquid and gaseous states. Solids differ from liquids and liquids from gases in the spacing and latitude of motion of their molecules, these variables being large in a gas, smaller in a liquid, and extremely small in a solid. Thus it follows that intermolecular Cohesive forces are large in a solid, smaller in a liquid, and extremely small in a gas

ics, with applications in all areas to the presentation of advanced theoretical or experimental research results. Modern computer packages are very good, but are of limited use unless the user is fully aware of their strengths and weaknesses. The student in has to decide whether the results of an analysis are physically meaningful. This requires a thorough understanding of the experiment being performed and a full appreciation of the reasons for plotting a graph at all.

Why plot graphs?. A clear picture reveals several things of data alone and provides answers to the following questions:

- i. How does a change in one variable lead to a change in the other?
- ii Do we have sufficient data?
- iii .Is there a region of interest that suggests further analysis?

Defining density

The density of an object is one of its most important and easily-measured physical properties. Densities are widely used to identify pure substances and to characterize and estimate the composition of many kinds of mixtures .

. These plots show how the masses of three liquids vary with their volumes. Notice that the plots all have the same origin of (0,0): if the mass is zero, so is the volume; the plots are all straight lines, which signify direct proportionality The only difference between these plots is their slopes. Denoting mass and volume by m and V respectively, we can write the equation of each line as $m = \rho V$, where the slope ρ (rho) is the proportionality constant that relates mass to volume. This quantity ρ is known as the **density**, which is usually defined as the mass per unit volume:

V.Manjula is currently working as an Assistant professor in Mathematics of Basic Engineering Department,

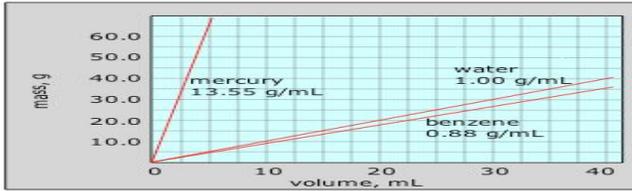
DVR& Dr. HS MIC College of Technology, Kanchikacherla A.P.,India

PH-9948233772,E-mail manju_adiraju@yahoo.co.in

2. IMPORTANCE

Drawing a graph is a natural part of doing Fluid Mechan-

$$\rho = m/V.$$



Specific volume

It is sometimes more convenient to express the volume occupied by a unit mass of a substance. This is just the inverse of the density and is known as the **specific volume**. **Specific gravity** is the ratio of the mass of a material to that of an equal volume of water. Because the density of water is about 1.00 g mL^{-1} , the specific gravity is numerically very close to that of the density, but being a ratio, it is dimensionless

Scope It is essentially a study of statics, kinematics and dynamics of fluid. It has important applications in diverse branches of mechanical, civil, Chemical Engineering. A Sound knowledge of fluid mechanics is essential the design of dams and irrigation structures fluid mechanics at present finds applications in new fields like bio-medical engineering, magneto-hydro dynamics. Being basic science it helps to develop an insight to common problems dealing with movement of air and water around us.

3..OBJECTIVES

Define the nature of a fluid.

- o Show where fluid mechanics concepts are common with those of solid mechanics and indicate some fundamental areas of difference.
 - o Introduce viscosity and show what are Newtonian and non-Newtonian fluids
 - o Define the appropriate physical properties and show how these allow differentiation between solids and fluids as well as between liquids and gases
2. There are two aspects of fluid mechanics which make it different to solid mechanics:
- o The nature of a fluid is much different to that of a solid

- o In fluids we usually deal with *continuous* streams of fluid without a beginning or end. In solids we only consider individual elements.

Properties of Fluids

The term fluid includes both liquid and gases. The main difference between a liquid and a gas is that the volume of a liquid remains definite because it takes the shape of the surface on or in which it comes into contact, whereas a gas occupies the complete space available in the container in which it is kept. In hydraulics in civil engineering, the fluid for consideration is liquid, so, we will examine some terms and properties of the liquids.

Viscosity: Viscosity is the property of fluid which defines the interaction between the moving particles of the fluid. It is the measure of resistance to the flow of fluids. The viscous force is due to the intermolecular forces acting in the fluid. The flow or rate of deformation of fluids under shear stress is different for different fluids due to the difference in viscosity. Fluids with high viscosity deform slowly.

Compressibility: When pressure is applied on a fluid, its volume decreases. This property of a fluid is called compressibility.

Elasticity: When the force generating the pressure on the fluid, is released it returns to its original volume. This property of a fluid is called elasticity of the fluid.

Vapor Pressure: Molecules of a liquid escape from its surface to fill the space above the liquid surface and the container until such time when the pressure due to these molecules above the liquid surface reaches the vapor pressure of the liquid. This is how the vapor pressure of a liquid is defined.

Surface Tension: The molecules on the surface of a liquid, that is, the interface between the liquid and the air are bound together by a weak force called surface tension. This force makes the liquid form a layer and is caused due to the cohesive force between the molecules of the liquid.

Capillarity: The molecules of a liquid have two types of forces acting on them. One is, cohesive force, the force among the molecules of the liquid only, and the other one is adhesive force, the force acting between the molecules of the liquid and some other substance. When the adhesion between the liquid and the container wall is more than the cohesion among the liquid molecules, the liquid sticks to the container walls and this results in capillary

rise. The opposite of this behavior happens when the cohesion is more than the adhesion - the capillary level dips

Viscosity in Liquids

There is some molecular interchange between adjacent layers in liquids but as the molecules are so much closer than in gasses the cohesive forces hold the molecules in place much more rigidly. This cohesion plays an important roll in the viscosity of liquids. Increasing the temperature of a fluid reduces the cohesive forces and increases the molecular interchange. Reducing cohesive forces reduces shear stress, while increasing molecular interchange increases shear stress. Because of this complex interrelation the effect of temperature on viscosity has something of the form: $\mu_T = \mu_0(1 + AT + BT^2)$ where μ_T is the viscosity at temperature TC, and μ_0 is the viscosity at temperature 0C. A and B are constants for a particular fluid. **Absolute Viscosity**

Absolute viscosity is the characteristic of a fluid which causes it to resist flow. The higher the numerical value of absolute viscosity assigned to a fluid, the greater the resistance that fluid offers to flow. The viscosity of a fluid causes a loss in pressure as it flows, so that an increase in viscosity requires an increased amount of energy to pump fluid at the same rate. When it is necessary to cause fluids to flow through small openings, such as in low capacity flow meters, too high a viscosity can cause so much pressure loss that it becomes impossible to establish the desired flow rate. Expressed another way, flow from a constant pressure source will decrease as the viscosity of the flowing fluid increases. The viscosity of a liquid is highly temperature dependent. An increase in temperature will cause a decrease in viscosity. For this reason, it is possible for a temperature change to affect the performance of a flowmeter considerably.

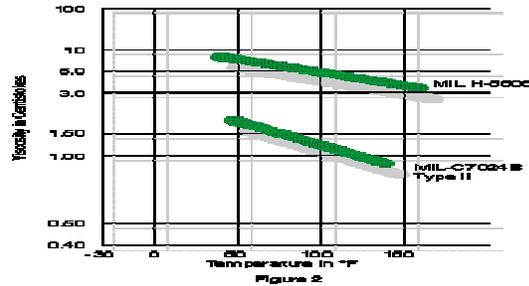
Absolute Viscosity

The ratio between shear forces and the velocity gradient in a fluid (as defined in the previous section) is absolute viscosity, and is identified with the symbol μ

$$\nu = \frac{\mu}{\rho}$$

Pressure The absolute viscosity of a fluid is strongly influenced by temperature. As temperature increases, the

viscosity of a liquid decreases and the viscosity of a gas increases. It is customary to express these relationships as a plot of viscosity vs. temperature, and such plots can be found in many references for common engineering fluids. Many oils have a straight line characteristic if the viscosity temperature relationship is plotted as on an ASTM chart



The influence of pressure on absolute viscosity is usually neglected, and this approximation is reasonable for low pressures. However, for pressures over about 1000 PSI, the absolute viscosity of a fluid may be a strong function of pressure. Generally, an increase in pressure will increase the viscosity of a liquid.

Kinematic viscosity is the ratio of absolute viscosity and density. Therefore, if density changes with temperature or pressure, the kinematic viscosity will also change. For gas applications, kinematic viscosity is a strong function of pressure.

When a viscous fluid flows over a solid surface, a force is exerted on the surface in a tangential direction. In effect, the moving fluid is attempting to drag the solid surface along with it. The magnitude of this force is dependent upon the viscosity and velocity of the fluid.

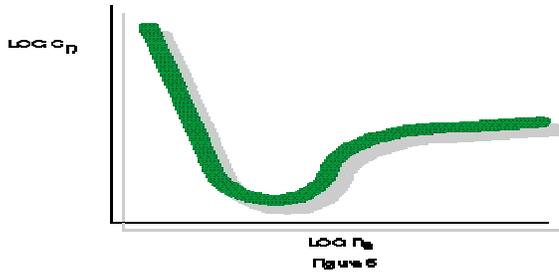
Experimentally measured drag forces are generally plotted in terms of CD vs. Re on a log-log chart. CD is the drag coefficient and Re is the Reynolds Number.

The drag coefficient is defined by:

$$C_D = \frac{F}{\frac{1}{2} \rho V^2 A}$$

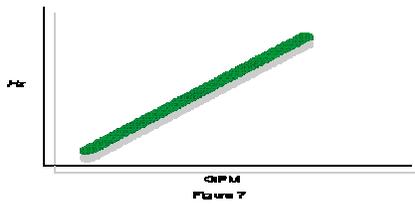
where
 F = Drag force
 V = Fluid velocity
 A = Wetted surface area

A plot of CD vs. Re usually looks something like Figure 6. The drag coefficient decreases rapidly with the Reynolds Number in laminar flow, rises abruptly in the transition region, then levels off and eventually decreases slowly in Turbulent region

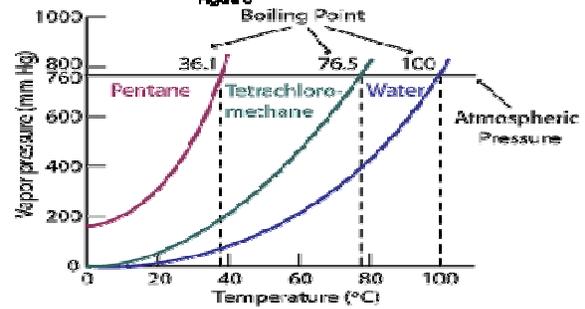
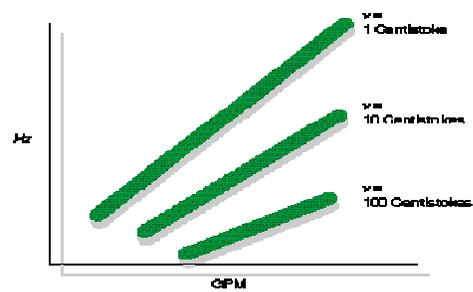


Because of the viscous retarding forces on the rotor, it does not spin as fast as it would in an in viscid fluid..The viscous drag also contributes to the pressure drop across the turbine meter.

For operation in high viscosity fluids, the curve in will have less slope and a positive zero offset along the horizontal axis.. Since a different curve will result for every viscosity, this is not a usable form for the calibration data except for single and constant viscosity operation.



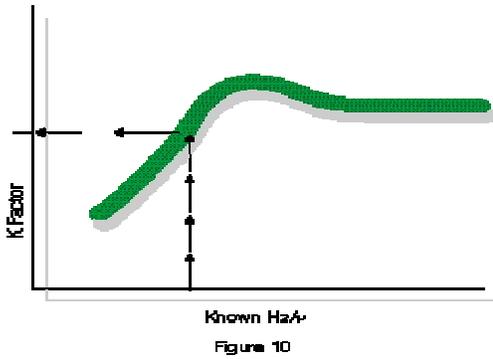
1. Determine output frequency Hz
2. Measure kinematic viscosity ν or measure temperature and use temperature to determine ν
3. Calculate Hz/ ν
4. Read up from known Hz/ ν to curve
5. Read over from curve to find K factor
6. Calculate GPM: $GPM = Hz \times 60/K$



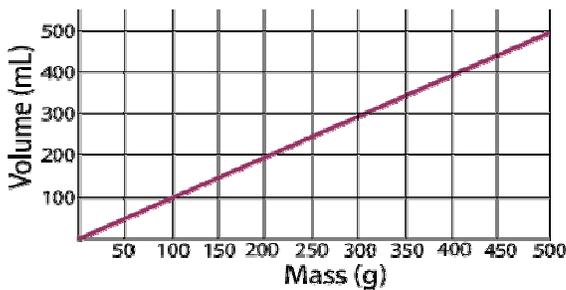
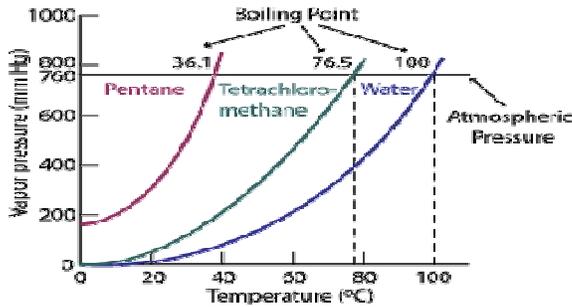
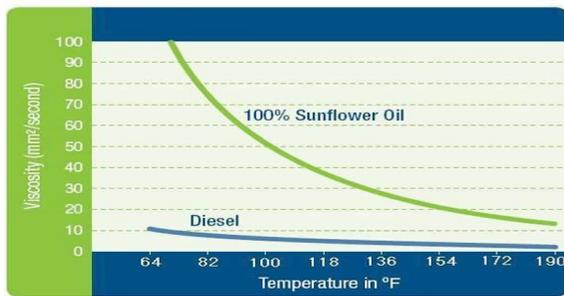
A more usable form for the calibration data is called a universal viscosity curve. This is a semi log plot of the sensitivity of the meter as a function of the ratio of the output frequency to the kinematic viscosity. the Universal Viscosity Curve is essentially a plot of meter sensitivity vs. Reynolds Number. As such, it reflects the combined effects of velocity, density and absolute viscosity acting on the meter. The latter two are combined into a single parameter by using kinematic viscosity (ν).

The Universal Viscosity Curve is formed by plotting K vs. HZ/ν for every calibration data point. Typically, thirty points are used; ten each for three different fluids. The thirty points are plotted on a common graph to form a smooth curve. Once this is done, the K factor may be determined for any flow rate in fluid of any viscosity as long as the ratio HZ/ν is within the range of values covered by the graph.

To determine the flow rate from measured output frequencies and viscosities simply follow the steps shown in Figure



4. GRAPHICAL REPRESENTATIONS



5. CONCLUSION

Thus Fluid has important applications in diverse branches of mechanical, civil, Chemical Engineering . A Sound knowledge of fluid mechanics is essential in the design of dams and irrigation structures a civil engineer can construct dams and pipelines across countries for transportation of oil, petrol, gases, etc. in these cases we should know the behavior r of fluids so that structures could be designed in such possible manner for ease of flow of fluids. fluid mechanics serves this purpose. Applications of fluid mechanics include a variety of machines, ranging from the water-wheel to the airplane

REFERENCES

- [1]. Dr. R.K.Jain]. Fluid Mechanics
- [2]. Dr.R.K.Bansal, Fluid Mechanics and Hydraulic Machines
- [3]. Narasimh DEO, Graph Theory and its Applications to Computer science
- [4] The importance of Graphs in Undergraduate Physics ,The Physics Teacher Vol. 37, May 1999
- [5] J.Mech Journal of Fluid Mechanics pg 153-163 ,2005

Data Mining Techniques for Real Time Intrusion Detection Systems

Monali Shetty, Prof. N.M.Shekokar

Abstract- Due to the widespread proliferation of computer networks, attacks on computer systems are increasing day by day. Preventive measures can stop these attacks to some extent, but they are not very effective due to various reasons. This lead to the development of intrusion detection as a second line of defense. In Information Security, intrusion detection is the act of detecting actions that attempt to compromise the confidentiality, integrity or availability of a resource. Intrusion detection does not, in general, include prevention of intrusions. In this paper, we are focused on data mining techniques that are being used for such purposes. We debate on the advantages and disadvantages of these techniques. Finally we present a new idea on how data mining can aid IDSs in real time.

In this paper, we present an overview of real time data mining-based intrusion detection system (IDSs). We focus on issues related to deploying a data mining -based IDS in a real time environment. New intelligent Intrusion Detection Systems (IDSs) are based on sophisticated algorithms rather than current signature-base detections are in demand. In this paper, we propose a new real time data-mining based technique for intrusion detection using an ensemble of binary classifiers with feature selection and multiboosting simultaneously.

Index Terms - Data Mining , DOS attack, Feature Selection ,Intrusion Detection Systems, Multiboosting, Network Security, Real time IDS

1. INTRODUCTION

A secure network must provide the following:

- Data confidentiality: Data that are being transferred through the network should be accessible only to those that have been properly authorized.
- Data integrity: Data should maintain their integrity from the moment they are transmitted to the moment they are actually received . No corruption or data loss is accepted either from random events or malicious activity.
- Data availability: The network should be resilient to Denial of Service attacks.

An intrusion detection system (IDS) inspects all inbound and outbound network activity and identifies suspicious patterns that may indicate a network or system attack from someone attempting to break into or compromise a system.

The rest of this work is a survey of data mining techniques that have been applied to IDSs and is organized as follows: In section 2 we debate on the drawbacks of standard IDSs. Section 3 offers a brief introduction to data mining and section 4 illustrates how data mining can be used to enhance IDSs. In section 5 we talk about the various data mining techniques that have been employed in IDSs by various researchers. Section 6 presents existing IDSs that use data mining techniques. In section 7, we give new proposal on how data mining can be used to aid IDSs, while in section 8 we conclude our work.

2. DRAWBACKS OF IDSS

Intrusion Detection Systems (IDS) have become a standard component in security infrastructures as they allow network administrators to detect policy violations. These policy violations range from external attackers trying to gain unauthorized access to insiders abusing their access.

Current IDS have a number of significant drawbacks:

- Current IDS are usually tuned to detect known service level network attacks. This leaves them vulnerable to original and novel malicious attacks.
- Data overload: Another aspect which does not relate directly to misuse detection but is extremely important is how much data an analyst can efficiently analyze. That amount of data he needs to look at seems to be growing rapidly. Depending on the intrusion detection tools employed by a company and its size there is the possibility for logs to reach millions of records per day.
- False positives: A common complaint is the amount of false positives an IDS will generate. A false positive occurs when normal attack is mistakenly classified as malicious and treated accordingly.
- False negatives: This is the case where an IDS does not generate an alert when an intrusion is actually taking place. (Classification of malicious traffic as normal)

Data mining can help improve intrusion detection by addressing each and every one of the above mentioned problems.

3. WHAT IS DATA MINING?

Data Mining, is the process of automatically searching large volumes of data for patterns using association rules. Here are a few specific things that data mining might contribute to an intrusion detection project:

- Find anomalous activity that uncovers a real attack
- Remove normal activity from alarm data to allow analysts to focus on real attacks
- Identify long, ongoing patterns (different IP address, same activity) To accomplish these tasks, data miners employ one or more of the following techniques:

- Data summarization with statistics, including finding outliers
- Visualization: presenting a graphical summary of the data
- Clustering of the data into natural categories
- Association rule discovery: defining normal activity and enabling the discovery of anomalies
- Classification: predicting the category to which a particular record belongs

4. DATA MINING AND IDS

The main function of the data mining model that we are interested in is classification, as normal, or malicious, or as a particular type of attack [10][12].

Common representations for data mining techniques include rules, decision trees, linear and non-linear functions (including neural nets), instance-based examples, and probability models [10].

4.1. Off Line Processing

The use of data mining techniques in IDSs, usually implies analysis of the collected data in an offline environment. There are important advantages in performing intrusion detection in an offline environment, in addition to the real-time detection tasks typically employed.

Below we present the most important of these advantages:

- In off-line analysis, it is assumed that all connections have already finished and, therefore, we can compute all the features and check the detection rules one by one [13].
- The estimation and detection process is generally very demanding and, therefore, the problem cannot be addressed in an online environment because of the various the real time constraints [16][15]. Many real-time IDSs will start to drop packets when flooded with data faster than they can process it.
- An offline environment provides the ability to transfer logs from remote sites to a central site for analysis during off peak times.

4.2. Data Mining and Real Time IDSs

Even though offline processing has a number of significant advantages, data mining techniques can also be used to enhance IDSs in real time. Lee et al. [17] were one of the first to address important and challenging issues of accuracy, efficiency, and usability of real-time IDSs. They implemented feature extraction and construction algorithms for labeled audit data.

.e.,g entropy, conditional entropy, relative entropy, information gain, and information cost to capture intrinsic characteristics of normal data and use such measures to guide the process of building and evaluating anomaly detection models. A serious limitation of their approaches (as well as with most existing IDSs) is that they only do intrusion detection at the network or system level.

However, with the rapid growth of e-Commerce and e-Government applications, there is an urgent need to do intrusion detection at the application-level. This is because many attacks may focus on applications that have no effect on the underlying network or system activities.

4.3. Sensor Correlation

The use of multiple sensors to collect data by various sources has been presented by numerous researchers as a way to increase the performance of an IDS.

- Kumar [12] states that, "Correlation of information from different sources has allowed additional information to be inferred that may be difficult to obtain directly."
- Lee et al.[17], state that using multiple sensors for ID should increase the accuracy of IDSs.

4.4. Evaluation Datasets

To test the effectiveness of data mining techniques in IDSs the use of established and appropriate datasets is required.

1. The DARPA datasets, available from the Lincoln Laboratory at MIT (<http://www.ll.mit.edu/IST/ideval>), are the most popular and widely used.

2. 'Knowledge Development and Data mining' (KDD) '99 cup challenge dataset.

These evaluations are contributing significantly to the intrusion detection research field by providing direction for research.

5. SURVEY OF APPLIED TECHNIQUES

In this section we present a survey of data mining techniques that have been applied to IDSs by various research groups.

5.1. Machine Learning

Machine Learning is the study of computer algorithms that improve automatically through experience. Applications range from data mining programs that discover general rules in large data sets, to information filtering systems that automatically learn users' interests. In contrast to statistical techniques, machine learning techniques are well suited to Clustering and Classification are probably the two most popular machine learning problems. Techniques that address both of these problems have been applied to IDSs.

5.1.1 Classification Techniques

In a classification task in machine learning, the task is to take each instance of a dataset and assign it to a particular class. A classification based IDS attempts to classify all traffic as either normal or malicious. The challenge in this is to minimize the number of false positives (classification of normal traffic as malicious) and false negatives (classification of malicious traffic as normal).

Five general categories of techniques have been tried to perform classification for intrusion detection purposes:

a) Neural Networks : The application of neural networks for IDSs has been investigated by a number of researchers. Neural networks provide a solution to the problem of modeling the users' behavior in anomaly detection because they do not require any explicit user model. Neural networks for intrusion detection were first introduced as an alternative to statistical techniques in the IDES intrusion detection expert system to model [29]. Numerous projects have used neural nets for intrusion detection using data from individual hosts [10].

McHugh et al [30] have pointed out that advanced research issues on IDSs should involve the use of pattern recognition and learning by example approaches for one reason:

- The capability of learning by example allows the system to detect new types of intrusion.

A different approach to anomaly detection based on neural networks is proposed by Lee et al. While previous works have addressed the anomaly detection problem by analyzing the audit records produced by the operating system, in this approach, anomalies are detected by looking at the usage of network protocols.

b) Fuzzy Logic : Fuzzy logic is derived from fuzzy set theory dealing with reasoning that is approximate rather than precisely deduced from classical predicate logic.

An enhancement of the fuzzy data mining approach has also been applied by Florez et al. [27] The authors use fuzzy data mining techniques to extract patterns that represent normal behavior for intrusion detection.

Luo [28] also attempted classification of the data using Fuzzy logic rules.

c) Genetic Algorithm : Genetic algorithms were originally introduced in the field of computational biology. Since then, they have been applied in various fields with promising results. Fairly recently, researchers have tried to integrate these algorithms with IDSs.

Chittur [25] applied a genetic algorithm and used a decision tree to represent the data. They used the "Detection rate minus the false positive rate" as their preference criterion to distinguish among the data. The REGAL System [23][24] is a concept learning.

d) Support Vector Machine : Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. SVMs attempt to separate data into multiple classes.

Mukkamala, Sung, et al. [32] used a more conventional SVM approach. They used five SVMs, one to identify normal traffic, and one to identify each of the four types of malicious activity in the KDD Cup dataset.

Eskin et al. [31], and Honig et al. [19] used an SVM in addition to their clustering methods for unsupervised learning. The achieved performance was comparable to or better than both of their clustering methods.

5.1.2 Clustering Techniques

Data clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics. Clustering is the classification of similar objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset (ideally) share some common trait - often proximity according to some defined distance measure. Machine learning typically regards data clustering as a form of unsupervised learning. Clustering is useful in intrusion detection as malicious activity should cluster together, separating itself from non-malicious activity. Clustering provides some significant advantages over the classification techniques already discussed, in that it does not require the use of a labeled data set for training.

Eskin et al. [31], and Chan et al. have applied fixed width and k-nearest neighbor clustering techniques to connection logs looking for outliers, which represent anomalies in the network traffic.

Bloedorn et al. [15] use a similar approach utilizing k-means clustering.

Marin et al. [34] employed a hybrid approach that begins with the application of expert rules to reduce the dimensionality of the data, followed by an initial clustering of the data and subsequent refinement of the cluster locations using a competitive network called Learning Vector Quantization. Since Learning Vector Quantization is a nearest neighbor classifier, they classified a new record presented to the network that lies outside a specified distance as a masquerader. Thus, this system does not require anomalous records to be included in the training set.

The authors were able to achieve classification rates, in some cases near 80% with misclassification rates less than 20%.

5.2. Feature Selection

"Feature selection, also known as subset selection or variable selection, is a process commonly used in machine learning, wherein a subset of the features available from the data is selected for application of a learning algorithm. Feature selection is necessary either because it is computationally infeasible to use all available features, or because of problems of estimation when limited data samples (but a large number of features) are present. Table I contains some examples of the features selected. Each of these features offers a valuable piece of information to the System.

TABLE 1.

FEATURES THAT HAVE BEEN EXTRACTED IN THE PROCESS OF APPLYING DATA MINING TECHNIQUES TO IDSS

Destination IP	# ICMP packets
----------------	----------------

Destination port	# to certain services
bytes transferred / all services	# total connections
Protocol	wrong data packet size rate
Source bytes	# urgent
TCP Flags	% data packet
Source port	# to privileged services
Duplicate ACK rate	# other errors
bytes transferred / current host	# packets to all services
% of same service to same host	# different services accessed
average duration / all services	# FIN flags

5.3. Statistical Techniques

Statistical techniques, also known as "top-down" learning, are employed when we have some idea as to the relationship we are looking for and can employ mathematics to aid our search.

Three basic classes of statistical techniques are linear, nonlinear (such as a regression-curve), and decision trees. Statistics also includes more complicated techniques, such as Markov models and Bayes estimators. Statistical patterns can be calculated with respect to different time windows, such as day of the week, day of the month, month of the year, etc. [33], or on a per-host, or per-service basis [13]. Denning (1987) described how to use statistical measures to detect anomalies, as well as some of the problems and their solutions in such an approach. The five statistical measures that she described were the operational model, the mean and standard deviation model, the multivariate model, the Markov process model, and the time series model.

Sinclair et al. [64] describe how they used Quinlan's ID3 algorithm to build a decision tree to classify network connection data. Bloedorn et al. [15] and Barbara et al. [65] also use decision tree-based methods.

1) Hidden Markov Models: Much work has been done or proposed involving Markovian models. For instance, the generalized Markov chain may improve the accuracy of detecting statistical anomalies. Unfortunately, it has been noted that these are complex and time consuming to construct [11], however their use may be more feasible in a high-power off-line environment. A hidden Markov model (HMM) is a statistical model where the system being modeled is assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters from the observable parameters. The extracted model parameters can then be used to perform further analysis, for example for pattern recognition applications.

A HMM can be considered as the simplest dynamic Bayesian network.

5.4. Ensemble Approaches

"In reality there are many different types of intrusions, and different detectors are needed to detect them." [Axelsson]

One way to improve certain properties, such as accuracy, of a data mining system is to use a multiplicity of techniques and correlate the results together. The combined use of numerous data mining methods is known as an ensemble approach, and the process of learning the correlation between these ensemble techniques is known by names such as learning, or meta-learning.

Lee and Stolfo [13][14][24] state that if one method or technique fails to detect an attack, then another should detect it. They propose the use of a mechanism that consists of multiple classifiers, in order to improve the effectiveness of the IDS.

6. EXISTING SYSTEMS

In this section, we present some of the implemented systems that apply data mining techniques in the field of Intrusion Detection.

1. The MINDS System [36]: The Minnesota Intrusion Detection System (MINDS), uses data mining techniques to automatically detect attacks against computer networks and systems. While the long-term objective of MINDS is to address all aspects of intrusion detection, the system currently focuses on two specific issues:

2. EMERALD (SRI) [35]: EMERALD is a software-based solution that utilizes lightweight sensors distributed over a network or series of networks for real-time detection of anomalous or suspicious activity. EMERALD sensors monitor activity both on host servers and network traffic streams. By using highly distributed surveillance and response monitors, EMERALD provides a wide range of information security coverage, real-time monitoring and response, protection of informational assets.

3. IDSs in the Open Market: Various systems that employ data mining techniques have already been released as parts of commercial security packages.- Dshield,, RealSecure SiteProtectort

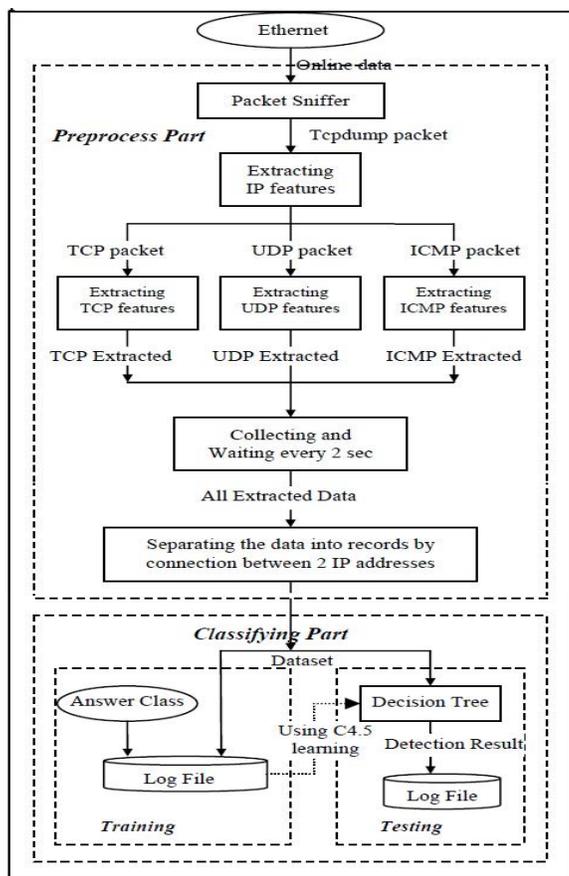


Fig. 2 Real-time IDS process(Existing System)

7. PROPOSED MODEL

In this section we propose a data mining technique that could potentially prove to be beneficial to Real Time IDSs. The idea is to use a new data-mining based technique for intrusion detection using an ensemble of binary classifiers with feature selection and multiboosting simultaneously.

We are making changes in Classifying Part .Our model employs feature selection so that the binary classifier for each type of attack can be more accurate, which improves the detection of attacks that occur less frequently in the training data. Based on the accurate binary classifiers, our model applies a new ensemble approach which aggregates each binary classifier’s decisions for the same input and decides which class is most suitable for a given input. During this process, the potential bias of certain binary classifier could be alleviated by other binary classifiers’ decision. Our model also makes use of multiboosting for reducing both variance and bias. In this model ,

For each trial $i, i=1 \dots T$, where T is the total no. of trials,
(1) A sample training set is generated by a multibooster using wagging (as specified in Webb’s multiboosting algorithm [15]).

(2) Binary classifiers are generated for each class of event using relevant features for the class and the C4.5

classification algorithm [13].

Binary classifiers are derived from the training sample by considering all classes other than the current class as other, e.g., Cnormal will consider two classes: normal and other. The purpose of this phase is to select different features for different classes by applying the information gain [18] or gain ratio [13] in order to identify relevant features for each binary classifier. Moreover, applying the information gain or gain ratio will return all the features that contain more information for separating the current class from all other classes. The output of this ensemble of binary classifiers will be decided using arbitration function based on the confidence level of the output of individual binary classifiers (e.g., see Fig. 2).

(3) The ensemble classifier is used by the multibooster in order to calculate the classification error, and derive the next training set.

(4) After T trials, the final committee is formed and it will be used by our intrusion detection system.

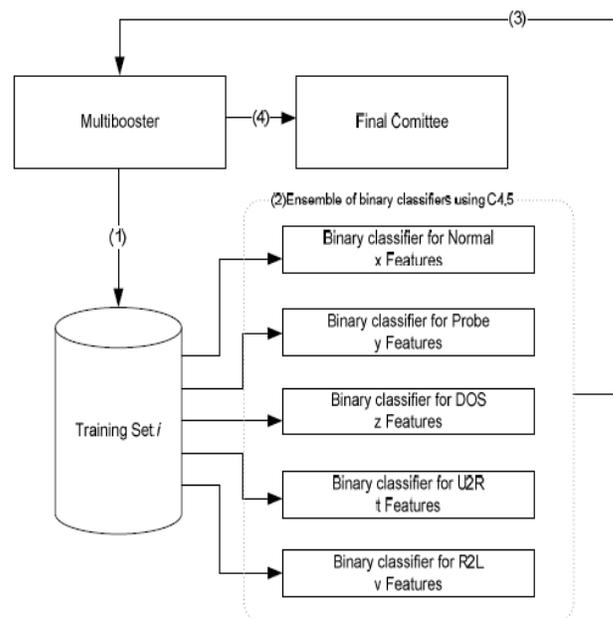


Fig. 3 The Diagram Of The Proposed Model

8. CONCLUSIONS

This paper has presented a survey of the various data mining techniques that have been proposed towards the enhancement of IDSs. We have shown the ways in which data mining has been known to implement the process of Intrusion Detection and the ways in which the various techniques have been applied and evaluated by researchers. Finally, in the last section, we proposed a data mining approach that we feel can contribute significantly in the attempt to create better and more effective Intrusion Detection Systems.

9. REFERENCES

- [1] Christos Douligeris, Aikaterini Mitrokotsa, "DDoS attacks and defense mechanisms: classification and state-of-the-art", *Computer Networks: The International Journal of Computer and Telecommunications Networking*, Vol. 44, Issue 5, pp: 643 - 666, 2004
- [2] Denning, D. E., "An intrusion-detection model", *IEEE Transactions on Software Engineering* 13 (2), 222-232, February, 1987.
- [3] <http://www.securityfocus.com/news/2445>
- [4] <http://www.webopedia.com>
- [5] Mithcel Rowton, Introduction to Network Security Intrusion Detection, December 2005
- [6] Biswanath Mukherjee, L. Todd Heberlein, Karl Levitt, "Network Intrusion Detection", *IEEE*, June 1994.
- [7] Presentation on Intrusion Detection Systems, Ariam Mavriqi
- [8] Intrusion Detection Methodologies Demystified Enterasys, Networks TM.
- [9] Fayyad, U. M., G. Piatesky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data," *Communications of the ACM* 39 (11), November 1996, 2734
- [10] Ghosh, A. K., A. Schwartzbard, and M. and Schatz, "Learning program behavior profiles for intrusion detection", In Proc. 1st USENIX, 9-12 April, 1999.
- [11] Kumar, S., "Classification and Detection of Computer Intrusion", PhD. thesis, 1995, Purdue Univ., West Lafayette, IN.
- [12] Lee, W. and S. J. Stolfo, "Data mining approaches for intrusion detection", In Proc. of the 7th USENIX Security Symp., San Antonio, TX. USENIX, 1998.
- [13] W. Lee, S. J. Stolfo et al, "A data mining and CIDF based approach for detecting novel and distributed intrusions", Third International Workshop on Recent Advances in Intrusion Detection (RAID 2000),
- [14] Lee, W., S. J. Stolfo, and K. W. Mok, "A data mining framework for building intrusion detection models," In 1999 IEEE Symp. On Security and Privacy, Oakland, CA, pp. 120132. IEEE Computer Society Press, 9-12 May 1999
- [15] Eric Bloedorn et al, "Data Mining for Network Intrusion Detection: How to Get Started," Technical paper, 2001.
- [16] Singh, S. and S. Kandula, "Argus - a distributed network-intrusion detection system," Undergraduate Thesis, Indian Institute of Technology, May 2001.
- [17] Lee, W. and D. Xiang, "Information-theoretic measures for anomaly detection", 2001 IEEE Symp. on Security and Privacy, IEEE 2001
- [18] Dickerson, J. E. and J. A. Dickerson, "Fuzzy network profiling for intrusion detection", NAFIPS 19th International Conference of the North American Fuzzy Information Processing Society, Atlanta, pp. 301306.
- [19] Honig, A., A. Howard, E. Eskin, and S. J. Stolfo, "Adaptive model generation: An architecture for the deployment of data mining based intrusion detection systems", In D. Barbar and S. Jajodia (Eds.), *Data Mining for Security Applications*. May 2002.
- [20] Helmer, G., J. Wong, V. Honavar, and L. Miller, "Automated discovery of concise predictive rules for intrusion detection", Technical Report 99-01, Iowa State Univ., Ames, IA, January, 1999.
- [21] Cohen, W. W., "Fast effective rule induction", Proc. of the 12th International Conference on Machine Learning, 9-12 July, 1995.
- [22] S. Stolfo, A. L. Prodromidis and P. K. Chan, "JAM: Java Agents for Meta-Learning over Distributed Databases", Third International Conference on Knowledge Discovery and Data Mining, 1997
- [23] Neri, F., "Comparing local search with respect to genetic evolution to detect intrusion in computer networks", In Proc. of the 2000 Congress on Evolutionary Computation CEC00, IEEE Press, 16-19 July, 2000.
- [24] Neri, F., "Mining TCP/IP traffic for network intrusion detection", *Machine Learning: ECML 2000*.
- [25] Chittur, A., "Model generation for an intrusion detection system using genetic algorithms", High School Honors Thesis, Ossining High School. In cooperation with Columbia Univ, 2001.
- [26] Theodoros Lappas and Konstantinos Pelechrinis, "Data Mining Techniques for Intrusion Detection Systems"
- [27] G. Florez, SM. Bridges, Vaughn RB, "An Improved Algorithm for Fuzzy Data Mining for Intrusion Detection", Annual Meeting of The North American Fuzzy Information Processing Society Proceedings, 2002.
- [28] Luo, J., "Integrating fuzzy logic with data mining methods for intrusion detection", Masters thesis, Mississippi State Univ., 1999.
- [29] Debar, H., Becker, M., and Siboni, D., "A Neural Network Component for an Intrusion Detection System", IEEE Computer Society Symposium on Research in Security and Privacy, Los Alamitos, CA, pp. 240-250, Oakland, CA, May 1992.
- [30] McHugh J., "Testing intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory", *ACM Trans. Information System Security* 3 (4), 262294, 2000.
- [31] Eskin, E., A. Arnold, M. Preraua, L. Portnoy, and S. J. Stolfo, "A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data", In D. Barbar and S. Jajodia (Eds.), *Data Mining for Security Applications*. Boston: Kluwer Academic Publishers, May 2002.
- [32] Mukkamala, S., A. H. Sung, "Identifying key variables for intrusion detection using neural networks", 15th International Conference on Computer Communications, pp. 1132-1138, 2002.
- [33] Frank, J., "Artificial intelligence and intrusion detection: Current and future directions", 17th National Computer Security Conference, Baltimore, MD. (NIST), 1994.
- [34] Marin, J. A., D. Ragsdale, and J. Surdu, "A hybrid approach to profile creation and intrusion detection", DARPA Information Survivability Conference and Exposition, Anaheim, CA. IEEE Computer Society, 12-14 June, 2001.
- [35] Porras, A. and Neumann, P. G., "EMERALD: Event Monitoring Enabling Responses to Anomalous Live Disturbances", National Information Systems Security Conference, October 1997.
- [36] Levent Ertöz and Eric Eilertson and Aleksandar Lazarevic and Pang-Ning Tan and Vipin Kumar and Jaideep Srivastava and

Paul Dokas, "MINDS - Minnesota Intrusion Detection System",
Next Generation Data Mining, MIT Press, 2004.

Aerothermal Energy

Jitul Moran

Abstract— The wind energy and solar energy are the future energy sources but both are not capitalized efficiently. This theory throws a light on the tremendous potential of both these energies when utilized in combination known as “AEROTHERMAL ENERGY”. This energy can produce a greater amount of power with minimum input. This energy also has the potential to produce power even in the absence of wind energy.

Index Terms— Aerothermal energy, Solar energy, Wind energy.

◆

1 INTRODUCTION

Currently wind energy and solar energy are the two forms of energy that are treated independently and they produce less energy compared to fossil fuel energy.

This experiment will test whether the solar and wind energy can be combined together to produce equivalent energy.

Winds are caused by the uneven heating of the atmosphere by the sun, the irregularities of the earth's surface, and rotation of the earth. The earth's surface is made of different types of land and water. These surfaces absorb the sun's heat at different rates, giving rise to the differences in temperature and subsequently to winds. During the day, the air above the land heats up more quickly than the air over water. The warm air over the land expands and rises, and the heavier, cooler air rushes in to take its place, creating winds. At night, the winds are reversed because the air cools more rapidly over land than over water. In the same way, the large atmospheric winds that circle the earth are created because the land near the earth's equator is heated more by the sun than the land near the North and South Poles. Humans use this wind flow for many purposes: sailing boats, pumping water, grinding mills and also generating electricity. Wind turbines convert the kinetic energy of the moving wind into electricity.

Wind Energy, like solar is a free energy resource. But is much intermittent than solar. Wind speeds may vary within minutes and affect the power generation and in cases of high speeds may result in overloading of generator. Energy from the wind can be tapped using turbines.

The range of wind speeds that are usable by a particular wind turbine for electricity generation is called productive wind speed. The power available from wind is proportional to cube of the wind's speed. So as the speed of the wind falls, the amount of energy that can be got from it falls very rapidly. On the other hand, as the wind speed rises, so the amount of energy in it rises very rapidly; very high wind speeds can overload a turbine. Productive wind speeds will range between 4 m/sec to 35 m/sec. The minimum prescribed speed for optimal performance of large scale wind farms is about 6 m/s. Wind power potential is mostly assessed assuming 1% of

land availability for wind farms required @12 ha/MW in sites having wind power density exceeding 200 W/sq.m. at 50 m hub-height.

In today's climate of growing energy needs and increasing environmental concern, alternatives to the use of non-renewable and polluting fossil fuels have to be investigated. One such alternative is solar energy.

Solar energy is quite simply the energy produced directly by the sun and collected elsewhere, normally the Earth. The sun creates its energy through a thermonuclear process that converts about 650,000,000 tons of hydrogen to helium every second. The process creates heat and electromagnetic radiation. The heat remains in the sun and is instrumental in maintaining the thermonuclear reaction. The electromagnetic radiation (including visible light, infra-red light, and ultra-violet radiation) streams out into space in all directions.

Only a very small fraction of the total radiation produced reaches the Earth. The radiation that does reach the Earth is the indirect source of nearly every type of energy used today. The exceptions are geothermal energy, and nuclear fission and fusion. Even fossil fuels owe their origins to the sun; they were once living plants and animals whose life was dependent upon the sun.

Much of the world's required energy can be supplied directly by solar power. More still can be provided indirectly. The practicality of doing so will be examined, as well as the benefits and drawbacks. In addition, the uses solar energy is currently applied to will be noted.

Due to the nature of solar energy, two components are required to have a functional solar energy generator. These two components are a collector and a storage unit. The collector simply collects the radiation that falls on it and converts a fraction of it to other forms of energy (either electricity and heat or heat alone). The storage unit is required because of the non-constant nature of solar energy; at certain times only a very small amount of radiation will be received. At night or during heavy cloudcover, for example, the amount of energy produced by the collector will be quite small. The storage unit can hold the excess energy produced during the periods of maxi-

mum productivity, and release it when the productivity drops. In practice, a backup power supply is usually added, too, for the situations when the amount of energy required is greater than both what is being produced and what is stored in the container.

Methods of collecting and storing solar energy vary depending on the uses planned for the solar generator. In general, there are three types of collectors and many forms of storage units.

People use energy for many things, but a few general tasks consume most of the energy. These tasks include transportation, heating, cooling, and the generation of electricity. Solar energy can be used for other things besides heating. It may seem strange, but one of the most common uses of solar energy today is cooling. Solar cooling is far more expensive than solar heating, so it is almost never seen in private homes. Solar energy is used to cool things by phase changing a liquid to gas through heat, and then forcing the gas into a lower pressure chamber. The temperature of a gas is related to the pressure containing it, and all other things being held equal, the same gas under a lower pressure will have a lower temperature. This cool gas will be used to absorb heat from the area of interest and then be forced into a region of higher pressure where the excess heat will be lost to the outside world. The net effect is that of a pump moving heat from one area into another, and the first is accordingly cooled.

Besides being used for heating and cooling, solar energy can be directly converted to electricity.

Of the main types of energy usage, the least suited to solar power is transportation. While large, relatively slow vehicles like ships could power themselves with large onboard solar panels, small constantly turning vehicles like cars could not. The only possible way a car could be completely solar powered would be through the use of battery that was charged by solar power at some stationary point and then later loaded into the car. Electric cars that are partially powered by solar energy are available now, but it is unlikely that solar power will provide the world's transportation costs in the near future.

Solar power has two big advantages over fossil fuels. The first is in the fact that it is renewable; it is never going to run out. The second is its effect on the environment.

While the burning of fossil fuels introduces many harmful pollutants into the atmosphere and contributes to environmental problems like global warming and acid rain, solar energy is completely non-polluting. While many acres of land must be destroyed to feed a fossil fuel energy plant its required fuel, the only land that must be destroyed for a solar energy plant is the land that it stands on. Indeed, if a solar energy system were incorporated into every business and dwelling, no land would have to be destroyed in the name of energy. This ability to decentralize solar energy is something that fossil fuel burning cannot match.

As the primary element of construction of solar panels, silicon, is the second most common element on the planet, there is very little environmental disturbance caused by the creation of solar panels. In fact, solar energy only causes environmental disruption if it is centralized and produced on a gigantic scale. Solar power certainly can be produced on a gigantic scale, too.

Suppose that of the 4.5×10^{17} kWh per annum that is used by the earth to evaporate water from the oceans we were to acquire just 0.1% or 4.5×10^{14} kWh per annum. Dividing by the hours in the year gives a continuous yield of 2.90×10^{10} kW. This would supply 2.4 kW to 12.1 billion people.

Unfortunately, at this scale, the production of solar energy would have some unpredictable negative environmental effects. If all the solar collectors were placed in one or just a few areas, they would probably have large effects on the local environment, and possibly have large effects on the world environment. Everything from changes in local rain conditions to another Ice Age has been predicted as a result of producing solar energy on this scale. The problem lies in the change of temperature and humidity near a solar panel; if the energy producing panels are kept non-centralized, they should not create the same local, mass temperature change that could have such bad effects on the environment.

2 METHOD

2.1 Apparatus

A turbine with alternate absorber and reflector porous metallic plate acting as blades.

2.2 Procedures: The turbine with alternate absorber and reflector porous metallic plate was exposed to the wind. The absorber and reflector plate were arranged in alternate fashion. The absorber plates were made black & the reflector metallic plates were made silver or shiny. Solar radiation was allowed to fall on all the plates.

2.3 Case Study:

Measurement of wind speed on DECEMBER 2003 MACHILIPATNAM CYCLONE

The variations in wind direction and wind speed are presented while the cyclone was crossing the coast. 1-hour average wind speed (in mps) as recorded by the system from 1600 h. of 15 Dec. to 1700 h. of 16 Dec. . . . The peak wind speed was observed between 1800 h. and 1900 h. Hence 10-minute average wind speeds (in mps) as recorded by the system between 1830 h. and 1930 h. 1-minute average wind speeds (in mps), as recorded by the system, between 1830 and 1930 h. 1-minute average wind speeds (in mps), as recorded by the system, when cyclone was 80km south of Machilipatnam . 1-minute average wind speeds (in mps), as recorded by the system, after the cyclone crossing the coast. Similarly, changes in wind direction (in degree), as observed before during and after the cyclone crossed the coast.

The wind speed/wind direction with time when cyclone was

about 80 km south of the station shows no significant variation. However, when cyclone was crossing the coast, following variations in wind speed and wind direction are observed: A sharp increase in wind speed from 14 mps to 27 mps between 1800 h. and 1900 h. and from 2000 h. and 2100 h. A sharp decrease in wind speed in between these two peaks, i.e. between 1900 h. and 2000 h. The peak hourly average wind speed recorded is of the order of 27 mps between 1800 h. and 1900 h. The peak 10-minute average wind speed recorded is of the order of 36 mps during 1900 h. and 1910 h. The peak 1-minute average wind speed recorded is of the order of 40 mps from 1847 h. and 1853 h. and again between 1902 h. and 1925 h. The change of wind direction from 'NNE' to 'southerly' is experienced between 1902 h. and 1930 h. The significant changes both in wind direction and wind speed is recorded between 1830 h. and 1930 h. of 15 Dec., 2003. The cyclonic wind experienced is seen to be for a very short duration and weakened immediately after crossing the coast. The wind speed tracked emphasizes the amount of wind speed available but not utilized to produce energy.

3 RESULT

From the above it can be observed that the wind speed varies and can produce large amount of energy if integrated with solar energy. The wind speed has kinetic energy and in combination with electromagnetic energy of the electromagnetic wave gives an energy known as aerothermal energy E_a .

Wind energy is the kinetic energy of air in motion;

Total wind energy flowing through an imaginary area A during the time t is:

$$E = 1/2mv^2 = 1/2(Avt \rho)v^2 = 1/2At\rho v^3 \quad (1)$$

where v is the wind speed; ρ is the air density; Avt is the volume of air passing ; $Avt\rho$ is therefore the mass m passing per unit time. $1/2 \rho v^2$ is the kinetic energy of the moving air per unit volume.

The energy of electromagnetic radiation is

$$E = hf \quad (2)$$

Where E is the energy, h is Planck's constant, and f is frequency

$$E_{rotational} = 1/2 I\omega^2 \quad (3)$$

Where

ω is the angular speed

I is the moment of inertia around the axis of rotation.

E is the kinetic energy.

$$E_a = j \{ (1/2At \rho v^3) + (hf) + (1/2 I\omega^2) \} \quad (4)$$

Where j =proportionality constant,
 E_a = aerothermal energy

After certain time in the absence of wind the equation reduces to:

$$E_a = j \{ (hf) + (1/2 I\omega^2) \} \quad (5)$$

as the absorbed radiation is radiated.

4 CONCLUSION

The turbine rotates as the metallic plates of the turbine experiences a force on it due the pressure difference created .The gas flows through the porous metallic plate from the cooler to the hotter side. The pressure difference causes the plate to move cold sided forward.

In the absence of wind the turbine turns backward because the black sides cool more quickly than the reflector sides.

Among the available renewable resources, only in aerothermal power do we find the potential for an energy source capable of supplying more energy than is used. This would be enough energy to supply the entire planet regardless of the population.

References

- [1] . Arnau J, Escobar JJ and Deltell A , "Wind energy in Catalonia trends. Environmental compatability" (Spain)
- [2] Bennett J, Planning for small wind turbines in the U.S.: A case study. (USA)
- [3] Asimov, Isaac; Understanding Physics: The Electron, Proton, and Neutron; Scarborough, Quebec: New American Library of Canada Ltd.; 1969
- [4] D.L.Elliott, C.G.Holladay, W.R.Barchet, H.P.Foote, W.F.Sandusky: Wind Energy Resource Atlas of the United States, Solar Energy Research Institute, Golden, Co, 1987
- [5] International Energy Agency (IEA) Wind Energy Annual Report 1998, National Renewable Energy Laboratory, Colorado, USA, April 1999
- [6] R.R. Mali, R.D. Vashistha & K.N. Mohan Monitoring of high wind speed by new state-of-the-art high wind speed recording system during recent December 2003 Machilipatnam cyclone

Significant Wavelet Hierarchical Approach Using Even Odd Method for Watermarking

S.Maruthuperumal, B.Vijayakumar, Dr.V.Vijayakumar

Abstract— Watermarking application requires flexibility to alter the cover image in order to embed the hidden information and makes the watermark more robust to attacks. These constraints are effectively achieved, in the present paper using SWH approach for digital watermark. In the proposed Significant Wavelet Hierarchical (SWH) approach, the selected subband of wavelet transform is divided in to non overlapped Hierarchical Regions (HR), each with sizes of $B \times B$ rows and columns. The HR is then divided into Significant Hierarchical Regions (SHR) of size $(B-i) \times (B-i)$ rows and columns. The SHR is used to represent the watermark. In the present approach watermark is inserted only in the SHR by a novel approach called REO Method. The unused row(s) and column(s) of HR are named as Unused Hierarchical Region (UHR). The UHR maintains the next level of hierarchy. This makes the present method as more efficient in terms of maintaining the hierarchy, security, authenticity, and in resisting various attacks. The experimental results with various attacks prove the efficacy of the proposed SWH approach.

Index Terms— Digital Watermarking, Discrete Wavelet Transformation, Significant Hierarchical Regions, Region based Even Odd

1 INTRODUCTION

Digital watermarking offers a means for protecting intellectual property of digital multimedia contents that have been explosively exchanged in the digital world. This technique is based on embedding information data (called watermark) into the digital contents. The main requirements of digital watermarking are invisibility, robustness and data capacity. These requirements are mutually conflicting, and thus, in the design of a watermarking system, the trade off has to be made [1].

Watermarking application requires flexibility to alter the cover image in order to embed the hidden information and makes the watermark more robust to attacks. These constraints are effectively achieved, in the present paper using SWH approach for digital watermark. In hierarchical digital watermarking methods the image is partitioned into non overlapping blocks, which constitute the lowest of the hierarchy, by combining distinct groups of blocks at a preceding level of the hierarchy [6]. In general, the number of blocks from a lower level of the hierarchy that are combined to form a block at the next level of the hierarchy may be arbitrarily chosen. Partitioning of the cover image into 'n' blocks is represented as given below in the figure 1.

Block based hierarchical watermarking system minimizes the changes in cover image when they are converted to corresponding watermark carrying regions in the watermarked image. Due to this nature, some regions in the cover object are

experiencing less change in the statistics after embedding. This process makes very difficult to identify the watermark, for an intruder.

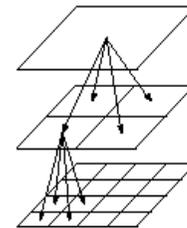


Fig.1 Hierarchical representation in the form of a quad-tree for a two dimensional image

The present paper is organized as follows. The section 2 describes the methodology, section 3 deals with the results and section 4 discusses regarding the conclusion of the paper.

2 SIGNIFICANT WAVELET HIERARCHICAL (SWH) APPROACH USING REGION BASED EVEN ODD (REO) METHOD

In the proposed Significant Wavelet Hierarchical (SWH) approach, the selected subband of wavelet transform is divided in to non overlapped Hierarchical Regions (HR), each with sizes of $B \times B$ rows and columns. The HR is then divided into Significant Hierarchical Regions (SHR) of size $(B-i) \times (B-i)$ rows and columns. The SHR is used to represent the watermark. In the present approach watermark is inserted only in the SHR by a novel approach called REO Method. By this $(B-(B-i))$ rows and columns are left unused for watermark insertion and they determine the address of the next hierarchical block where the watermark is inserted i.e., next level

- S.Maruthuperumal is currently pursuing Ph.D in Computer Science and Engineering in Dr.M.G.R.University, Chennai,India
E-mail: maruthumail@gmail.com
- B.Vijayakumar is currently pursuing Ph.D in Computer Science and Engineering in GITAMS University,Vizag, India, E-mail: vijaysree.b@gmail.com
- Dr.V.Vijayakumar is currently working as Dean Computer Sciences & Head Srinivasa Ramanujan Research Forum-GIET Rajahmundry,India, E-mail: vakulabharanam@hotmail.com

of hierarchy. The unused row(s) and column(s) are named as Unused Hierarchical Region (UHR). In the present paper, SHR are chosen from $(B-2) \times (B-2)$ rows and columns leaving top most and bottom most rows and left most and right most columns as unused for watermark. One bit from three pixels which constitute three bits are used to indicate the next level of hierarchy. The same address is preserved in the top most, bottom most, left most and right most column of the UHR. This phenomenon sustains any type of attack. The next level of hierarchy is taken at the time of extraction if and only if all four or three UHR show the same address. This makes the present method as more efficient in terms of maintaining the hierarchy, security, authenticity, and in resisting various attacks.

To maintain high security in the proposed hierarchical system the next hierarchical block address is given in the sixth bit position of 2nd, 3rd and 4th pixels of the top row of UHR. The same is also represented in the same positions of the bottom most left and right most columns of UHR.

2.1 Embedding Algorithm for SWH approach using REO method

The entire algorithm of the proposed Significant Wavelet Hierarchical (SWH) based approach using the REO Method is given below.

Algorithm: SWH approach using REO method of digital watermarking

Begin

Step 1: Apply n level DWT on the cover image and obtain the nth level LL subband image.

Step 2: Divide the nth level LL subband image into non overlapped hierarchical blocks of size $B \times B$ rows and columns.

Step 3: Divide the Hierarchical block into the SHR with a window or sub block of size $(B-2) \times (B-2)$ rows and columns.

Step 4: Arrange the gray level values of SHR in ascending order along with their coordinate positions, $P_i(x_i, y_i)$, $P_{i+1}(x_{i+1}, y_{i+1})$; here $P_i(x_i, y_i)$ denotes the gray level value of the location (x_i, y_i) .

Step 5: Consider successive even (e_i) and odd gray values (e_{i+1}) as same after sorting. Where $((e_{i+1})-e_i)$ is always one and $e_i < e_{i+1}$. If two or more pixels of the SHR have the same gray level value or if they are successive even and odd values of the SHR then the least coordinated value of row and column is treated as least value. The watermark bit is embedded in the ascending order of gray level values of the considered SHR on the least x coordinate and y coordinate position.

Step 6: Convert each character of the watermark in to a 12 bit character by appending the MOD 9 value of each character.

Step 7: Insert the bits of watermark in to the identified pairs in ascending order of step 4.

Step 8: Place the next hierarchical block address from the nth bit position of three successive pixels of the UHR.

Step 9: If the watermark insertion process is over place the next level of hierarchical address as zero (000) and go to the step 10. Otherwise repeat the process in the next hierarchical block from step 3.

Step 10: Stop.

End of the algorithm

In the proposed SWH using REO method, successive

even and odd values are treated as same but not successive odd and even values. Because an even number will have always a zero in the LSB, even by embedding a '1' in the LSB, its value is incremented by one at most. In the same way an odd number is always having a one in the LSB, even by embedding a '0' in the LSB its value is at most decremented by one. i.e., the odd values will never increment by 1 after embedding the digital watermark bit. And the even values will never decrement by 1 after embedding the digital watermark bit.

Therefore the maximum difference between successive even and odd values will be always one after embedding the digital watermark bit. Whereas the maximum difference between successive odd and even values is two after inserting the digital watermark bit. For this reason, the successive even and odd values of a neighborhood are treated as same in the proposed approach. This property removes the ambiguity in the extraction process of the watermark bits, based on ascending order of the window. The entire algorithm of the proposed SWH based approach using the REO method is explained with an example given below.

Consider that the DWT subband is divided into 7×7 bit non overlapping hierarchical blocks. Then on each hierarchical block, a SHR is formed with 5×5 bits leaving the top and bottom most rows and left and right most columns as shown in figure 2. The figure 2 shows a 7×7 non overlapped HR with gray level values. In figure 2 the UHR are shown with red border and the three pixels position of UHR which are used for storing the next level of hierarchy is shown in green. In the present experiment, the next level of hierarchy is considered from the sixth bit position of 2nd, 3rd and 4th pixel locations of UHR, and this is shown in green color.

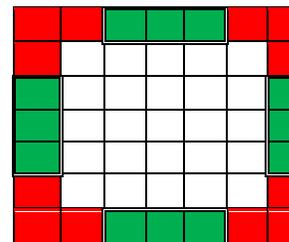


Fig. 2 Non overlapped HR with 7×7 bits

In the proposed method the first block has chosen as the first HR and the next level of hierarchy is determined by the 3 bit address of the UHR.

2.2 Extraction algorithm

The DWT is applied on the watermarked image. The watermarked image in the DWT is divided into non overlapped hierarchical blocks of size $B \times B$ rows and columns. The SHR is obtained by dividing the HR in to $(B-2) \times (B-2)$ rows and columns by leaving the top and bottom most rows and left most and right most columns. The watermark is obtained from LSB positions of SWH approach using REO method. For the next level of hierarchical address, the 3 bit address of top most and bottom most row and left and right most columns of the UHR are compared. If three or all four UHR show the same address then next HR address is chosen and

the process is repeated. Otherwise the process is stopped saying that an attack has occurred and changed the hierarchical block address, which is a rare phenomenon.

3 RESULTS AND DISCUSSION

The proposed SWH method using REO method is experimented on 24 images of size 512×512 with the text “srinivasaramanujan”. In the present paper Haar wavelet is used in the SWH method. The few watermarked images and the extracted images are given in figure 3 and figure 4 respectively. Table 1 shows the PSNR and NCC values for all the considered 24 images. From the table 1 it is clearly evident that all the images shows high PSNR value above 60 dB and NCC value above 0.92. The average PSNR value for the considered images is 64.23 dB and NCC is 0.94 without attacks. This indicates high robustness and high quality of the images after watermark insertion.

TABLE 1 QUALITY MEASURES OF THE WATERMARKED IMAGES FOR THE PROPOSED SWH APPROACH USING REO METHOD

S.No	Original images	PSNR	NCC
1	Lena	64.76	0.94
2	House	65.25	0.92
3	F16	64.47	0.94
4	Joker	65.38	0.94
5	Living room	64.23	0.93
6	Child	66.37	0.94
7	Monalisa	64.25	0.93
8	Milkdrop	64.09	0.94
9	Baboon	63.54	0.94
10	Bear	63.23	0.94
11	Fruit	65.12	0.93
12	Line	64.48	0.93
13	Circle	64.49	0.99
14	Brain	60.37	0.94
15	Barbara	65.32	0.93
16	Brick	63.46	0.99
17	Pepper	65.64	0.93
18	Flight	63.39	0.94
19	Lake	64.78	0.93
20	Flower	64.29	0.97
21	MRI Scan	63.57	0.96
22	Cameraman	63.34	0.99
23	Boat	63.45	0.95
24	Lungs	64.37	0.96
	Average	64.23	0.94

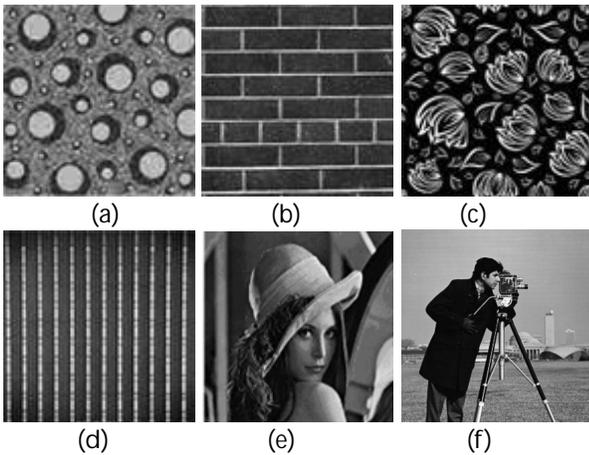


Fig.3 Watermarked images (a) Circle (b) Brick (c) Flower (d) Line (e) Lena (f) Cameraman

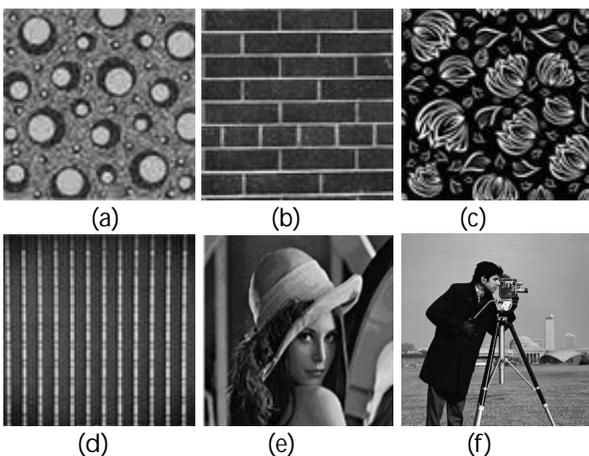


Fig.4 Extracted cover images (a) Circle (b) Brick (c) Flower (d) Line (e) Lena (f) Cameraman

Table 2 clearly indicates that the robustness and quality of the image is not degraded for all attacks.

The proposed SWH method using Haar wavelets is tested with various attacks such as Cropping with different ratios (2%, 3% and 4%), Rotation with 2, 3 and 4 degrees, Gaussian noise with different ratios (3%, 5% and 7%) and Median filter with different size (3×3, 5×5 and 7×7), to test the robustness, and shown in table 2.

TABLE 2 RESULTS OF THE EXPERIMENTS ON SWH METHOD WITH VARIOUS ATTACKS ON THE WATERMARKED TEXT

S.No	Attacks	Circle		Brick		Flower	
		PSNR	NCC	PSNR	NCC	PSNR	NCC
1	Cropping (2%)	59.32	0.85	59.24	0.83	59.28	0.81
2	Cropping (3%)	58.01	0.71	58.68	0.79	58.82	0.70
3	Cropping (4%)	57.16	0.73	57.43	0.66	57.97	0.71
4	Rotation 2°	61.14	0.80	60.61	0.84	60.86	0.79
5	Rotation 3°	60.02	0.70	59.42	0.77	58.63	0.69
6	Rotation 4°	57.41	0.68	57.53	0.69	56.01	0.67
7	Gaussian noise (3%)	62.36	0.87	62.86	0.81	61.11	0.86
8	Gaussian noise (5%)	61.12	0.81	60.51	0.75	59.74	0.80
9	Gaussian noise (7%)	59.16	0.78	58.41	0.71	57.69	0.77
10	Filtering (3×3)	61.39	0.86	62.36	0.82	62.32	0.85
11	Filtering (5×5)	59.28	0.81	60.41	0.77	59.56	0.80
12	Filtering (7×7)	58.23	0.78	59.33	0.71	57.67	0.77

3.1 Comparison of the proposed SWH using REO method with various other methods

Table 3 compares the PSNR values after inserting the watermark without attacks with various other existing methods [2, 3, 4 and 5]. Table 2 clearly indicates that SWH with REO method outperforms the other existing methods. The remaining three methods show a PSNR value between 29.42 dB and 51.48 dB, where as the proposed method shows a minimum of 63.34 dB and an average of 64.13 dB.

TABLE 3 COMPARISON OF SWH APPROACH USING REO METHOD WITH VARIOUS OTHER METHODS

Test Images	Wei-Hung Lin et.al method	Wei-Hung Lin et.al method	Mei Jian-sheng et.al method	V. Padmanabha Reddy et.al method	Proposed SWH using REO method
	PSNR(dB)				
Circle	42.14	42.45	50.12	29.42	64.49
Brick	41.89	42.48	49.34	29.79	63.46
Flower	42.02	41.98	50.04	29.67	64.29
Line	42.72	42.81	51.34	29.85	64.48
Lena	42.02	42.53	50.02	29.54	64.76
Cameraman	42.61	42.53	51.48	29.56	63.34

Table 3 clearly indicates that the proposed method shows very high PSNR value when compared with the other existing methods.

4 CONCLUSIONS

The present paper proposed an effective wavelet based hierarchical watermarking scheme with Haar wavelets for protecting copy rights of digital images.

The proposed SWH approach created a new direction for the feature researchers in hierarchical watermarking methods, by dividing the each hierarchical block into SHR and UHR. The proposed SWH approach using REO watermarking system minimizes the changes in cover image when they are converted to corresponding watermark carrying regions in the watermarked image. Due to this nature, some regions in the cover object will experience less change after embedding. This process makes it very difficult to identify the watermark, for an intruder. That's why the present scheme can be used for both copyright protection and for encryption. The novelty of the proposed REO method is, it embeds the information in a non linear order based on the values and position of a window. Table 2 clearly indicates the proposed two steps SWH with REO watermarking scheme outperforms the other existing methods [2, 3, 4 and 5]. The proposed SWH method exhibited a high PSNR and NCC value, with various attacks such as Cropping, Rotation with different degrees, Gaussian noise and Median filter with different size to test the robustness.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Sri K.V.V. Satyanarayana Raju, Founder & Chairman, and Sri K. Sasi Kiran Varma, Managing Director, Chaitanya group of Institutions for providing necessary Infrastructure. Authors would like to thank the anonymous reviewers for their valuable comments.

REFERENCES

- [1] S.-H. Wang and Y.-P. Lin, "Wavelet tree quantization for copyright protection watermarking", IEEE Trans. Image Processing, vol. 13, no. 2, pp. 154-165, Feb, 2004.
- [2] Wei-Hung Lin, Yuh-Rau Wang and Shi-Jinn Horng (2009), "A Block-Based Watermarking Method Using Wavelet Coefficient Quantization," CA3PP 2009, Springer-Verlag Berlin Heidelberg, pp. 156-164.
- [3] Wei-Hung Lin, Yuh-Rau Wang, Shi-Jinn Horng, Tzong-Wann Kao, Yi Pan (2009), "A blind watermarking method using maximum wavelet coefficient quantization," Elsevier Ltd.
- [4] Mei Jiansheng, Li Sukang and Tan Xiaomei (2009), "A Digital Watermarking Algorithm Based On DCT and DWT," International Symposium on Web Information Systems and Applications, pp. 104-107.
- [5] Padmanabha Reddy.V and Dr. Varadarajan .S (2010), "An Effective Wavelet-Based Watermarking Scheme Using Human Visual System for Protecting Copyrights of Digital Images," International Journal of Computer and Electrical Engineering, Vol. 2, No. 1, February.
- [6] Celik .M.U, Sharma. G, Saber. E, Tekalp . A.M (2002), "Hierarchical watermarking for secure image authentication with localization," IEEE Trans. Image Process. 11 (6).

Analysis of Internet Security Measures Surrounding the World

Amit Kumar Jain*, Yashpal Singh, K.K Pandey, Sachin Upadhyay

Abstract The objective of this paper is to estimate the statistics surrounding the most common security threats faced by Internet users. There is an estimated of more than two billion Internet users worldwide, therefore it is important to know what security threats your computer may be vulnerable to while using the Internet. Threats discussed in this paper will include spam, phishing, computer viruses, hackers, and spyware/malware. The current percentage of Incidents as they are related to different regions of the world discusses the severity of each threat, by using suitable statistical techniques. Due to the large number of Internet users, it is probable that many of them are unaware of these threats and what they can and should be doing to protect themselves. Most importantly this paper will discuss about threats. A user can take to defend themselves against these threats and known vulnerabilities. With identity theft on the rise, it is imperative to understand Internet security threats now more than ever.

Keywords: Computer Security, Trojan horse, Spyware, etc.

1 INTRODUCTION

There are many security Measures that face computers in the world today, and we are going to see at a few of them as they relate to the Internet. Since its inception, the Internet has grown from original purpose as a military tool to a worldwide phenomenon. According to the latest statistical analysis, it is estimated that more than two billion internet users are worldwide. The following table provides the statistical breakdown of world internet usage.

The Internet is full of useful information, in fact, it is estimated that there are between 15 and 30 billion different websites in existence today. Considering this estimate of available websites, it is easy to see that the Internet is an invaluable resource to many people.

*Author for correspondence.

Table-1: World Internet Users and Population Statistics

World Regions	Population (2011 Est.)	Internet Users Dec 31, 2010	Internet users latest data	Penetration(% population)	Growth 2000 - 2011	Users% of table
Africa	1037524058	4514400	118609620	11.4%	2527.4%	5.7%
Asia	3879740877	114304000	922329554	23.8%	706.9%	44.0%
Europe	816426346	105096093	476213935	58.3%	353.1%	22.7%
Middle East	216258843	3284800	68553666	31.7%	1987.0%	3.3%
North Amer	347394870	108096800	272066000	78.3%	151.7%	13.0%

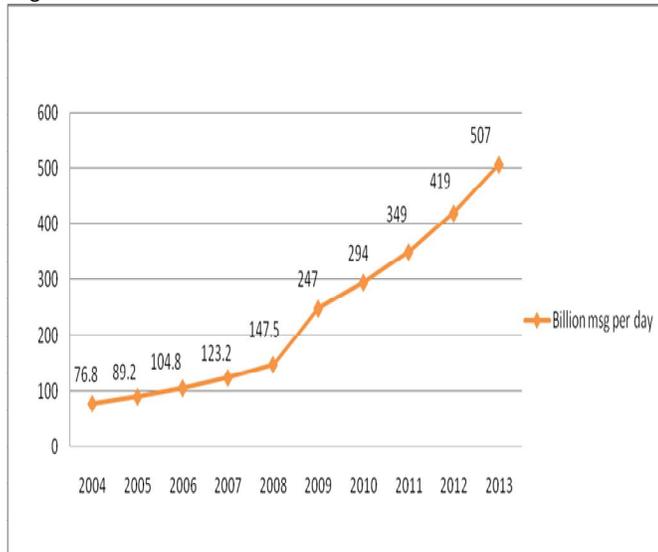
Region	Population	Internet Users	Penetration	Growth	Users%
Latin America	597283165	18068919	215939400	36.2%	1037.4%
Oceania/Australia	35426995	7620480	21293830	60.1%	179.4%
World Total	6930055154	360985492	2095006005	30.2%	480.4%

The Internet provides many diverse and useful resources such as email, instant messaging, academic research, product research, paying bills, shopping, online banking, and the list goes on and on. For many of the Internets 2.09 billion users the Internet is not just a tool but a way of life. Businesses and people all over the world rely heavily on the Internet to perform their vital daily tasks (Table-1). The Internet has become such an integral part of global society to the extent that the world would be hard pressed to continue forward with such great progress without it. There are so many well known advantages to using the Internet, however many users fail to take the time to research the risks involved. It is important to know the risks involved in any activity we decide to pursue in life and the Internet is no exception. The risks associated with the Internet are realized in the form of information security threats or vulnerabilities. The risks discussed in this paper include spam, phishing, Trojan viruses, hackers, and spyware/malware. This paper will also discuss some measures you, as a user, can take to help secure yourself and your computer against these Internet security threats.

Description of Project

Email is a very useful tool that many people use daily in their personal business endeavors. According to Radicati, 1.4 billion people around the world now use email regularly. This figures expected to grow steadily over the next two years, reaching 1.9 billion users by the end of 2013.

Figure-1: Email Traffic, 2004-2013



Internet Email Traffic Worldwide

Email is very convenient, but with that convenience comes several security risks. The most common and potentially the most harmful email security threat is not in what you send but what is sent to you. Junk email, or Internet solicitations, is a huge security risk. This type of email is widely known by the name of spam. Time wasted deleting junk e-mail costs American businesses more than \$30 billion a year. Sending an email to someone is the virtual equivalent of sending someone a postcard through regular post office mail. For this reason, it is a good idea to use encryption when sending an email that contains confidential information. A Telephone-based survey of adults who use the Internet found that more than 75% receive spam daily. The average spam messages per day are 20.5, and the average time spent per day deleting them is 2.9 minutes. The loss in productivity is equivalent to \$25 billion per year at average US wages, according to the National Technology Readiness Survey produced by Rockbridge Associates and the Center for Excellence in Service at Maryland's business school. 14% of spam recipients actually read messages to see what they say, and 4% of the recipients have bought

Something advertised through spam within the past year. The best defense against spam is to use a spam filter. A computer user needs to be aware of what spam is

and is not so they can make informed decisions when an email arrives in their inbox. If you use Outlook 2010 or higher there is a built-in spam filters that you can configure to your personal requirements. It is also needful on a corporate or enterprise level to use a hardware spam filter to block known spam before it reaches the end users. This will save you much time and money later and is worth the investment. While it is important to defend against spam, it is nearly impossible to filter it all out. This is why user education is so important.

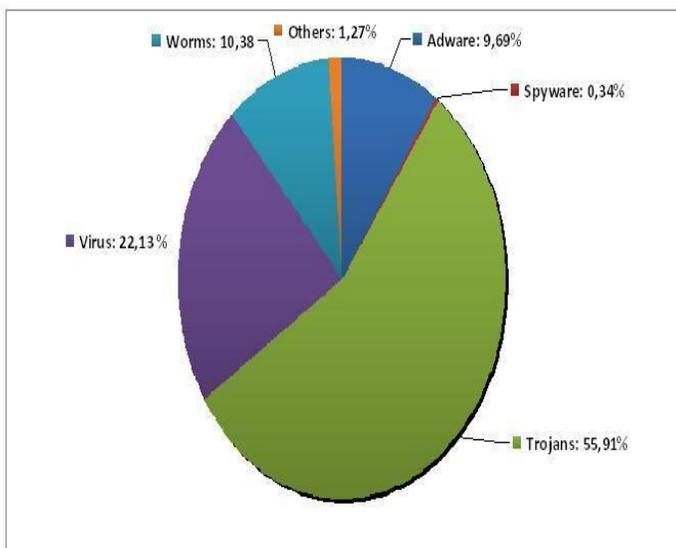
Email users are also being targeted by a different type of spamming technique called phishing. Phishing is a fraudulent attempt, usually made through email, to steal your personal information. The best way to protect you from phishing is to learn how to recognize a phish. A phishing email attempt will appear to many users to be a legitimate email perhaps from a reputable company or bank. However, the intent of the sender is to tempt you into giving them your personal information such as your social security number, usernames and passwords, and even your bank account or credit card numbers. This is done by sending huge amounts of spam phishing emails to many users by someone claiming for example to be your bank. The phishing email may state that your bank account information needs to be updated and will provide a hyperlink to a website that looks like your bank's website.

However, this is not your bank's website, but one created by the phisher to look just like it! You use your login information, and update your personal information and logout thinking you have updated your information, but what you have really done is given your information to a thief. The phisher will then use your personal information to steal your identity and your money. You can defend yourself against phishing attempts by being aware of procedures. A bank will never send you an email asking you for your personal information. Most of the banks correspondence will be done with post office mail or with a phone call. It is vitally important to investigate any email or link to a website you receive via email before you input any of your personal information. Microsoft's Internet Explorer 7 actually has a built in anti-phishing filter that will scan websites against a pool of known phishing sites. While this is not fool proof, it is an added defense against phishing attempts. This feature must be turned on to work, and this can be accomplished through Internet options under tools on the file menu. Again, user education and an awareness of procedure is the best defense against this type of threat or scam.

Another common Internet security measure is becoming infected with a computer virus. A computer

virus can be passed many ways such as via email, floppy disk, CDRW, flash drive, network connection, or a hacker breaking into your system. There are many different computer viruses in existence today. Each one is different and their creators had different motives or functions for the virus to perform. There were over 50,000 computer viruses in 2000 and that number was then and still is growing rapidly. Sophos, in a print ad in June 2005 claims "over 103,000 viruses." And, Symantec, in April 2008 is reported to have claimed the number is over one million. Fortunately, only a small percentage of these are circulating widely. Some viruses will simply cause your data to become corrupt, while others are designed to steal your data or create a backdoor into your system via the Internet, which are called Trojan's.

Figure-2 Annual security report 2010

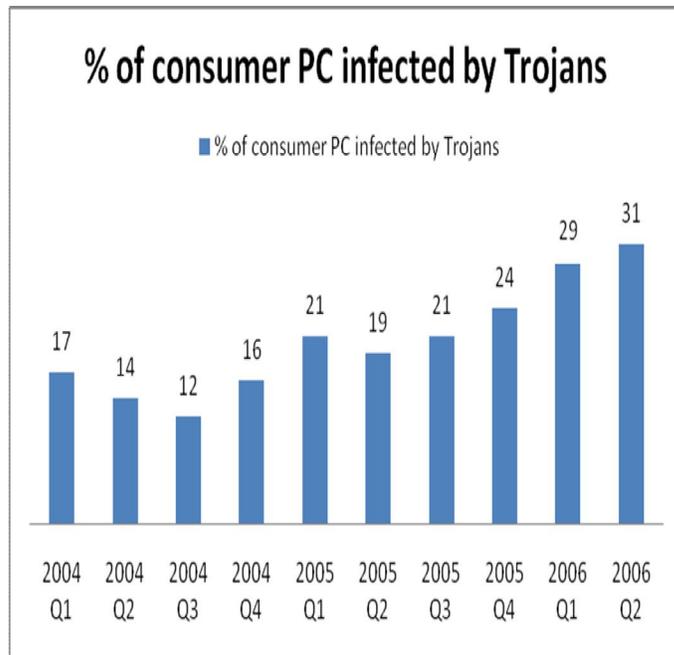


Trojans still dominate the ranking of new malware that has appeared in 2010 (56 percent of all samples), followed by viruses and worms. It is interesting to note that 11.6 percent of all the malware gathered in the Collective Intelligence database is rogware or fake antivirus software, a malware category that despite appearing only four years ago is creating much havoc among users.

Every day viruses cause a huge amount of data loss, and in turn cost individuals time and money. The best defense against computer viruses is to install an antivirus program on every computer you own. An antivirus program can only detect a virus if it knows the virus exists, and it does this via virus definitions. Since new viruses are constantly being created it is Imperative to keep your antivirus definitions up to date and by using a package

with an automatic update feature will do this for you. Also, be sure the antivirus you use utilizes real-time protection, which will quickly identify the presence of a virus. There are many different antivirus vendors, and there are equally as many opinions on which one are the best to use. When selecting an antivirus product, make sure it includes an automatic update feature.

Figure-3 The infection of Trojan analyze by webroot is given below from 2004 to 2006



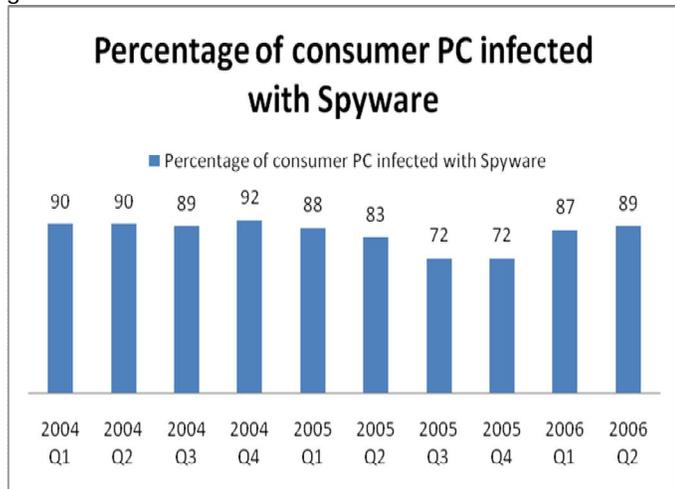
Trojan Infections from 2004 – mid 2006

It is also important that your antivirus program scans email attachments automatically for viruses. Since many viruses are transmitted via email this can be a valuable tool! First and foremost, it is important as a user to be educated and aware of potentially harmful files. Never open any files or emails if you do not know the person that sent them to you. Following this rule can save you a lot of trouble later.

Another growing security threat is something known as spyware. Spyware is a type of malware that can be installed on computers, and which collects small pieces of information about users without their knowledge. If you notice your computer is abnormally slow all of sudden, receives many pop-up advertisements, or your homepage has been hijacked, your computer is likely infected with spyware. Here are three shocking statistics reported by PCSecurityNews.com, 8 out of 10 PC's are infected with

some sort of Spyware, with an average of 24.4 spies per PC scanned, Microsoft estimates that 50% of all PC crashes are due to spyware, Dell reports that 20% of all technical support calls involve spyware.

Figure-4 The infection of Spyware analyze by webroot is given below from 2004 to 2006



Spyware Infections from 2004 – mid 2006

The presence of spyware is typically hidden from the user, and can be difficult to detect. Typically, spyware is secretly installed on the user's personal computer. Sometimes, however, spywares such as key loggers are installed by the owner of a shared, corporate, or public computer on purpose in order to secretly monitor other users. When you look at these statistics it is easy to see that spyware is a very real threat to all PC's connected to the Internet, and many users are unaware that they are victims of spyware.

There are several defenses against spyware. To help stop the spread of spyware and other malware, it is essential to be alert to suspicious activity on your computer and to learn safe computing practices. While some spyware is deployed by exploiting flaws in operating systems or applications, much of it still relies on social engineering to trick you into running or installing malware. You must exercise caution when downloading anything from public web sites, newsgroups, instant messaging sessions, or when opening email attachments, even from senders you know. Identity is often difficult to verify on the internet. Frequently, attackers and their malware impersonate associates of the target user to coax them into installing the malicious code. A common example of this is when malware infects a system and then automatically emails itself to everyone in the infected person's address book. When such an email is received, the recipient is more likely

to open the contents because the sender is a familiar, trusted source.

Don't trust unknown or known high-risk sources when visiting unfamiliar web sites, you should exercise caution. This guideline should also apply to sites you expect to be high risk based on their content. Such sites include those with many popup.

Pay attention when installing applications Software installation packages sometimes take advantage of a user's tendency to not pay attention to the details and simply accept the default "checked" options. If the default options are blindly accepted and prompts are ignored, clicking next, next, next may actually be agreeing to the installation of spyware, adware, or other applications that are not desired. Reading instructions and paying attention to what is being agreed to is important to staying safe. Keep your operating system and software up to date keeping systems and applications current with security-related patches is critical. This includes patching the operating system and all installed applications, especially those related to network and internet activity like browsers, media players, email clients, and news readers. These are very common targets of attack and second only to social engineering as a means of spreading malware.

Installing trusted antivirus and antispyware tools and keeping them and their signatures current is an important part of defensive computer security. There are many packages available for purchase and some available for free to download, such as Spybot and Ad-Aware. Microsoft has even joined the fight against spyware with their free for download program called Windows Defender. One of the best defenses against spyware is to prevent infection by developing safe Internet surfing habits. In other words, stay away from questionable websites. Spyware not only comes from websites but you can also be infected by Peer to Peer file sharing. Spyware and Viruses run rampant on P2P file sharing networks such as Lime Wire, Kazaa, Bear share, Gnutella, Grokster, and eDonkey. When you connect to these and other P2P networks to share files, the chances are you do not know who you are downloading the file from or who is downloading files from you. Forty-five percent of the executable files downloaded through Kazaa contain malicious code. It is the best practice not to use these types of services as a spyware or virus infection is likely to occur on your computer.

Another Internet security threat is hacking. Computer hacking is the practice of modifying computer hardware and software to accomplish a goal outside of the creator's original purpose. People who engage in computer hacking activities are often called hackers. Since the word "hack" has long been used to describe someone who is incompetent at his/her profession, some hackers claim this term is offensive and fails to give appropriate recognition to their skills. While it remains a very interesting subject or hobby for computer techies, it is a very serious threat and should not be taken lightly. A hacker may attempt to access your computer or network for a number of reasons, which include file storage, information for identity theft, malicious intent, or even just for fun. Many computers and networks have been compromised by hackers around the world, and the users are unaware they have been hacked. The best defense against hacking is to setup a strong defense perimeter.

A good basic defense should consist of a firewall, strong passwords (at least 8 characters long utilizing both numeric, alphanumeric, and special characters), the latest software patches for your operating system and applications, and Antivirus/Antispyware software with updated definitions. PSINet Europe purposely built an unprotected server and connected it to the Internet to determine how quickly it would be compromised. Their findings were astonishing: the server was maliciously attacked 467 times in the first 24 hours, most of the attacks originated in the US or Western Europe, and after 3 weeks a total of 626 attacks were detected against the server [8]. It is easy to see from this case study project that if you have a computer connected to the Internet without proper security, it will be compromised very quickly. It is especially important for users with a broadband Internet connection to maintain security due to the nature of the "always on" Internet connection. In this case your computer is always vulnerable to attack while it is powered on unless you have the network connection disabled or unplugged.

CONCLUSION

The goal of this paper is to help those users understand the Seriousness of current Internet security threats and to show them ways to protect their personal information. After compiling and analyzing these Internet security threat statistics, the only possible conclusion is that the Internet, while very useful, is not to be taken lightly.

Due to the commercialization and ease of use of the Internet in the last decade, it is only reasonable to

conclude that the Internet will grow as society becomes more reliant on it and its conveniences. Every Internet user should be aware and educated of the threats and vulnerabilities that surround the Internet and know what to do to protect themselves against these known threats. Internet users should be encouraged to stay abreast of current threats and defense mechanisms by using the Internet itself as a research tool. There are many good sources on the Internet for current and past threats and how to setup a defense against them. The irony is that you can use the Internet to learn how to make your Internet surfing more secure. It is always important to know the risks of any activity a person chooses to pursue in life, and the Internet is no exception. It also never hurts to get a knowledgeable friend or consultant to take a look at your current configuration and make suggestions on how to harden your security. In conclusion, the Internet is full of useful material but this comes at a risk. It is important to develop safe surfing habits and a strong security plan before connecting to and utilizing the Internet.

With this conclusion, it is also reasonable to conclude that new Internet security threats will likely arise in the coming months and years, and therefore will require users to become even more proactive in defending their computer systems.

REFERENCES

- 1- World Internet Users and Population Stats. (2011, March 31). Internet World Stats. Retrieved March 31, 2011 from the WWW: <http://www.internetworldstats.com/stats.htm>
- 2- Sara Radicati, Email Statistics Report, 2009-2013; the radicati group, inc. a technology market research firm, Palo Alto, CA.
- 3- The report on malware to find out Trojan. (<http://press.pandasecurity.com/news/2010-annual-security-report>)
- 4- The size of the World Wide Web. (2007, February 25). Pandia Search Engine News. Retrieved March 20, 2007 from the WWW: <http://www.pandia.com/sew/383-web-size.html>
- 5- Security Statistics. (2005) Aladdin: Securing the Global Village. Retrieved March 21, 2007. (WWW: http://www.esafe.com/home/csrt/statistics/statistics_2005.asp)

- 6- Some Interesting RSA Phishing Stats. (2006, November 9) ZDNet.co.uk. Retrieved March 21, 2007 from the WWW:
<http://community.zdnet.co.uk/blog/0,1000000567,10004498o-2000331828b,00.htm>

Dynamic Filtering Points Based Skyline Query Processing In Distributed Data Sites

M.Amuthavalli, Prof. P.S.Balamurugan

Abstract— The Skyline query used in a large number of unstructured distributed databases located at different data sites, it extracts the data from, where relevant data are distributed among geographically scattered sites. Constrained skyline query is attached with constraints on specific dimensions. A constraint on a dimension is a range specifying the user's interest. Given a skyline query with constraints, all relevant sites are partitioned into incomparable groups, among which the query can be executed in parallel based on partition algorithm. We then give a parallel distributed skyline algorithm to estimate the overall query response time. We have used some filtering points in distributed query processing such that the amount of data to be transmitted via the network connection is reduced. In this paper we propose dynamic filtering points to reduce the large number of filtering points used in a same data sites and get a good accurate result. In previous we used skyline only in centralized approach, now we are used skyline in distributed environment at larger network. The results of an extensive experimental study demonstrate that our proposals are effective and efficient.

Index Terms— Dynamic filtering points, partition algorithm, parallel distributed skyline algorithm



1 INTRODUCTION

Skyline is used in distributed database, because the database will not be in one system. It will be stored in multiple systems at different locations, if it is connected using internet. A Query is called as Skyline which query works or execute based on data points. Skyline query returns many multidimensional points. It extracts information from different places of distributed database at different sites. Skyline query returns all interesting points that are not dominated by any other points. Skyline queries play an important role in multi criteria decision making and user preference applications. For instance, a tourist can issue a skyline query on a hotel relation to get those hotels with high stars and cheap prices.

Most works on skyline queries so far have assumed a centralized data storage, and been focused on providing efficient skyline computation algorithms on a sole database. This assumption, however, fails to reflect the distributed computing environments consisting of different computers, which are located at geographically scattered sites and connected via Internet. Given a distributed environment without any overlay structures, our objective is efficient query processing strategies that shorten the overall query response time.

We first speed up the overall query processing by achieving parallelism of distributed query execution. Given a skyline query with constraints, all relevant sites are partitioned into incomparable groups among which the query can be executed in parallel. The parallel execution

also makes it possible to report skyline points progressively, which is usually desirable to users. Within each group, specific plans are proposed to further improve the query processing involving all intra group sites.

The filtering points are used to communicate between the query and distributed sites. On a processing site, dynamic filtering points are calculatingly picked based on their overall dominating potential from the local skyline. They are then sent to other sites with the query request, where they help identify more unqualified points that would otherwise be reported as false positives, and thus, reducing the communication between data sites. We propose a specific partition algorithm that helps to extract information quickly by partition then it divides all relevant sites into groups such that a given query can be executed in parallel among all those site groups. Then it is making interrelated sites as a block. We elaborate on the intra group query execution strategies. We then give a parallel distributed skyline algorithm to estimate the overall query response time and it work parallel at different sites at a particular time. We detail heuristics for selecting a dynamic filtering point in distributed query processing such that the amount of data to be transmitted via the network is reduced. .

The filtering points which is created to remove unwanted data from skyline query database. In previous paper they propose a cost-efficient model for dynamically determining the number of filtering points to be sent to a particular site

such that the benefit of using filtering points is maximized. In this paper we propose dynamic filtering points to use generate an efficient query in short time. We conduct an extensive experimental study on real data sets, and the results demonstrate the effectiveness, efficiency, and robustness of our proposals.

2 PROJECT OVERVIEW

In this project we use five modules to get a accurate result from distributed data sites and reduce the query response time.

2.1 SKYLINE QUERY WITH CONSTRAINTS

This is the first module of project which is concern with the process of forming skyline query. The module first need the dataset as input this input is processed and various rules are taken it consideration and these rules are identified. Based on condition of these rules the constraints are identified.

2.2 CONSTRAINTS FETCH DATASITES

The distributed database systems in each database are located as different places each place will be distributed. This distributed database is called as a dataset. From this data sites the skyline query would be executed. Each dataset identified by based on the database located at different distributed data places where a data sites are virtually found in finding the results for skyline query.

2.3 FORMATION OF SITES AND CONSTRAINTS

After identifying the constraints and sites the next important thing is to fit the constraints to the dataset and fetch the details in which a result will be skyline query. A formation of sites is based on different location and form of constraints is based on condition.

2.4 RESULT

After formation of conditions and data sites the is generated the results are based on the various points that is calculated point in order to filtering points for different skyline. Each data sites have multiple filtering points which based upon the skyline query. The filtering point review is the result is taken as group of results.

2.5 SELECT THE DATA

Once the result is got the next step is to identify and fetch the results of the query. The query results are got from the execution of the skyline query. The execution of the skyline query based on the data sites and the constraints give set of data that uses the select skyline query and fetches the result.

3 ABOUT THE DATASET

Skyline query processing is with respect to the process of distributed. A distributed database has data set interact which has fields of students that is taken from the fields of student name, college, place, percentage, sex. These fields play a important role in finding the result of queries. A dataset fields has relevant to a process of skyline query.

4 BACKGROUND STUDY AND RELATED WORK

Data Mining is a multidisciplinary field, drawing work from areas including database management systems, artificial intelligence, machine learning, neural networks, statistics, pattern recognition, knowledge-based systems, knowledge acquisition, information retrieval, high-performance computing, and data visualization.

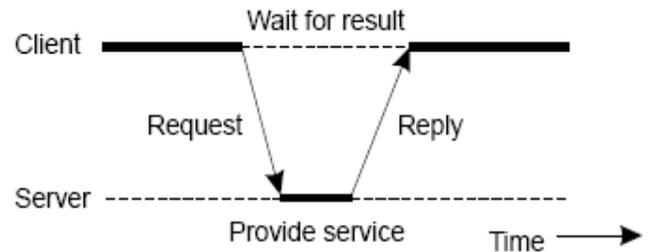


Fig.1 Basic Client-Server Model in distributed systems

Data mining extracts the information's from large distributed data bases it increase the query response time we introduce the skyline to distributed systems to reduce the overall query response time.

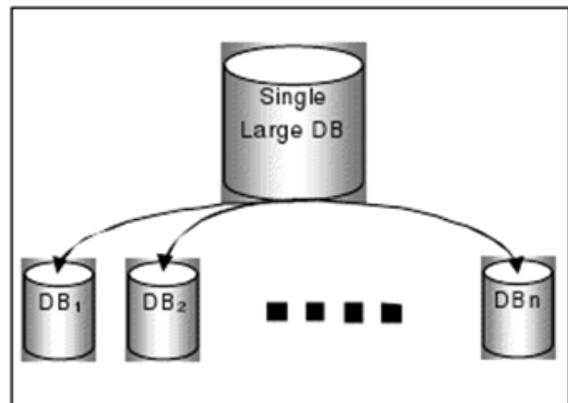


Fig.2 Partitioning of Database

5 PARALLEL DISTRIBUTED QUERY EXECUTION

Query execution order among different sites, the query with condition first asks each site S_i for its MBR $_i$, then-dimensional minimum bounding box of the local relation R_i . If a site's MBR disjoins with the constraint set C specified in the query, it will not be considered in the following query processing. For each site whose MBR $_i$ overlaps with C , we only need to consider the intersection $MBR_i \cap C$. We call this intersection reduced minimum bounding box and use $rMBR_i$ to represent it. We proceed to partition all those sites left into several groups according to their $rMBR$ s such that the skyline computation in any one group does not depend on or affect the computation in any other group. Therefore, the given skyline query can be executed in parallel among those site groups. While within any individual site group, the query is executed according to some local plans, which will be Partition algorithm. A distributed database is not stored in its entirety at a single physical location. Instead, it is spread across a network of computers that are geographically dispersed and connected via communications links.

6 PROBLEM DEFINITION

Given a set of N sites $S_1; S_2; \dots; S_n$ distributed at different geographic locations, each S_i has a local relation R_i . Every tuple in any R_i is an n -dimensional point, represented as (p_1, p_2, \dots, p_n) . Different R_i s may overlap; it is possible that $R_i \cap R_j$. Without loss of generality, we assume that smaller values are preferred in the skyline operator. We use $pt_1 \succ pt_2$ to represent point pt_1 dominates point pt_2 . In addition, we suppose any site S_{org} , able to directly communicate with any other site $S_i \in S$ through wired end-to-end connections, may initiate against all R_i a skyline query with a set of n constraints C_1, C_2, \dots, C_n . Each C_i is either a range $[l_i; u_i]$, or a indicating no constraint in that dimension. Our goal is to get the result for the constrained skyline query efficiently, i.e., with short response time. We define the query response time as the time period from the moment a query is issued by a site S_{org} to the moment S_{org} receives the complete and correct answers after contacting other sites. To shorten the response time of a query, we mainly endeavor on two aspects. On one hand, we propose effective ways to guide deciding the query forwarding and execution order between different sites, so as to obtain the query execution parallelism, and boost the result reporting progressiveness. On the other hand, we generalize the single filtering point idea to use dynamic filtering points, and thus, enhancing the filtering power and reducing the amount of data transmitted between remote sites. In contrast to previous

work, their problem definition does not assume the availability of an overlay network, where different nodes hold disjoint data partitions. Instead, different relations on different sites may overlap. Therefore, their previous methods are not applicable to our problem.

7 QUERY PROCESSING IN DISTRIBUTED DATABASE

Query processing in a distributed database system requires the transmission of data between computers in a network. The arrangement of data transmissions and local data processing is known as a distribution strategy for a query. Two cost measures, response time and total time are used to judge the quality of a distribution strategy. Simple algorithms are presented that derive distribution strategies which have minimal response time and minimal total time, for a special class of queries.

Today's data is rarely stored in centralized location due to the enormous amount of information that needs to be stored and also to increase reliability, availability and performance of the system. Same data is stored in different format into different company's database as well as they may be partitioned or replicated. We consider various scenarios of distributed database such as horizontal, vertical fragmentation and attribute overlapping. Allowing access to integrated information from these multiple datasets can provide accurate and wholesome information to the end-user. We research on efficient querying to these distributed databases to get top k elements matching the ranking order provided by the user. We also discuss hierarchical way of using the top k algorithm and their limitations to our problem. In today's world of universal dependence on information systems, all sorts of people need access to companies' databases. In addition to a company's own employees, these include the company's customers, potential customers, suppliers, and vendors of all types. It is possible for a company to have all of its databases concentrated at one mainframe computer site with worldwide access to this site provided by telecommunications networks, including the Internet. Although the management of such a centralized system and its databases can be controlled in a well-contained manner and this can be advantageous, it poses some problems as well.

- Parallelized search
- Irrelevant nodes pruning
- Reduction of duplicate query forwarding

8 SKYLINE QUERY PROCESSING WITH RESPECT TO THE DATASET

Skyline query processing executed based on the data set. These data sets are stored in relevant distributed databases in different sites. First we give the skyline query with constraints all relevant sites are partitioned into incomparable groups among which the query can be executed in parallel.

The parallel execution also makes it possible to report skyline points progressively, which is usually desirable to users. Within each group, specific plans are proposed to further improve the query processing involving all intra group sites.

The filtering points are used to communicate between the query and distributed sites. On a processing site, dynamic filtering points are calculatingly picked based on their overall dominating potential from the local skyline. They are then sent to other sites with the query request, where they help identify more unqualified points that would otherwise be reported as false positives, and thus, reducing the communication between data sites.

We propose a specific partition algorithm that helps to extract information quickly by partition then it divides all relevant sites into groups such that a given query can be executed in parallel among all those site groups. Then it was making interrelated sites as a block. We elaborate on the intra group query execution strategies.

We then give a parallel distributed skyline algorithm to estimate the overall query response time and it work parallel at different sites at a particular time. We detail heuristics for selecting dynamic filtering points in distributed query processing such that the amount of data to be transmitted via the network is reduced. The filtering points which is created to remove unwanted data from skyline query database. In this paper we propose a dynamic filtering point to use generate a efficient query in short time. We conduct an extensive experimental study on both synthetic and real data sets, and the results demonstrate the effectiveness, efficiency, and robustness of our proposals.

9 CONCLUSION

In previous paper they used feed-back-based skyline algorithm for horizontal processing and multiple filtering points in distributed skyline query processing in different distributed data sites. It increases the query response time and transmission of network bandwidth size.

In this paper, we have addressed the problem of constrained skyline query processing against distributed data sites. To accelerate the query processing, we partition

all relevant sites into incomparable groups and parallelize the query processing among all groups.

We select local skyline points and send them as dynamic filtering points together with the query to relevant data sites, in order to prevent more data from being transmitted through the network. Furthermore, a dynamic filtering point's selection strategy is proposed based on a PaDSkyline (parallel distributed skyline algorithm) to reduce the query response time and get effective query.

Extensive experimental results demonstrate the efficiency and effectiveness of our proposals in a distributed network environment.

REFERENCES

- [1] W.-T. Balke, U. Guntzer, and J.X. Zheng, "Efficient Distributed Sky lining for Web Information Systems," Proc. Int'l Conf. Extending Database Technology (EDBT), pp. 256-273, 2004.
- [2] P. Wu, C. Zhang, Y. Feng, B.Y. Zhao, D. Agrawal, and A.E. Abbadi, "Parallelizing Skyline Queries for Scalable Distribution," Proc. Int'l Conf. Extending Database Technology (EDBT), pp. 112-130, 2006.
- [3] S. Wang, B.C. Ooi, A.K.H. Tung, and L. Xu, "Efficient Skyline Query Processing on Peer-to-Peer Networks," Proc. IEEE Int'l Conf. Data Eng. (ICDE), pp. 1126-1135, 2007.
- [4] Y. Lu, J. Zhao, L. Chen, B. Cui, and D. Yang, "Effective Skyline Cardinality Estimation on Data Streams," Proc. Int'l Conf. Database and Expert Systems Applications (DEXA), pp. 241-254, 2008
- [5] Y. Tao, L. Ding, X. Lin, and J. Pei, "Distance-Based Representative Skyline," Proc. Int'l Conf. Data Eng. (ICDE), pp. 892-903, 2009.
- [6] L. Zhu, Y. Tao, and S. Zhou, "Distributed Skyline Retrieval with Low Bandwidth Consumption," IEEE Trans. Knowledge and Data Eng., vol. 21, no. 3, pp. 384-400, Mar. 2009.



M. Amuthavalli, Received BE degree in 2010 from Anna University of technology Coimbatore and currently pursuing ME Degree at Karpagam University, Coimbatore



Prof. P.S. Balamurugan received BE degree in 2003 from Bharathiyar University Coimbatore and ME in 2005 from Anna University. Currently he is pursuing his Doctoral degree at Anna University, Coimbatore.

Minimization of Leakage Current in VLSI Design

Kanika Kaur, Arti Noor

Abstract - To meet the ever-increasing demand of high performance systems, more and more functions are integrated into single chip by scaling down the size of device. Leakage current is becoming an increasingly important fraction of total power dissipation of integrated circuits. As technology scales leakage current grows exponentially and become an increasingly large component of total power dissipation. An important area of research is developing the circuit techniques to reduce the subthreshold leakage current in both active and standby mode to minimize the total power consumption. The power dissipation due to subthreshold leakage current becomes comparable to switching or dynamic power component and is a serious concern for circuit designing in deep submicron region. This paper describes the need to consider gate leakage current while determining the sleep state pattern is explained. Circuit reorganization and sleep state assignment techniques demonstrated for gate and subthreshold minimization of static and dynamic circuits. The MTCMOS technology for the minimization of gate and subthreshold leakage current is also explained for the low power circuits.

Index Terms: - Gate Leakage, Subthreshold leakage Current, Static circuits, Dynamic circuits, MTCMOS, CMOS

◆

I. INTRODUCTION

In the past, the figure of merits for VLSI Design was high speed, low cost and small area without much bothering about power dissipation but in present scenario demand of low power design is addressed. In recent years, the aggressive scaling of device dimensions and threshold voltage have significantly increased sub-threshold leakage and its contribution to the total chip power consumption. Also, gate oxide thickness has been scaled to maintain adequate control of the channel by the gate. This has resulted in an alarming increase of late in gate leakage current due to tunneling through the thin gate oxide. Gate leakage is expected to be a major component of leakage in future technology generations and has been identified as one of the most important challenges to future device scaling [1]. Gate leakage power, which was almost non-existent in the previous technology generations, is expected to contribute more than 15% to the total chip power dissipation in the today's technology generations. To date, most circuit-level leakage minimization techniques focus only on sub-threshold leakage reduction, without considering the effects of gate leakage. Gate leakage is primarily being addressed from a CMOS technology perspective and the use of high-k gate dielectrics have being proposed. One of the approaches that addresses gate leakage, BGMOS [2], uses multiple threshold voltages and multiple oxide thickness devices. The use of PMOS dominated circuits was proposed in [3], on the basis that PMOS devices exhibit lower gate leakage compared to identical NMOS devices. However, due to band-to-band tunneling and use of different dielectrics [4], the gate leakage through PMOS devices is no longer negligible and needs to be considered. In this paper, we account for the contribution of gate leakage on total leakage by considering forward and reverse gate tunneling through both

NMOS and PMOS devices. Gate leakage in conventional sleep-state patterns (which focus only on sub-threshold leakage) are evaluated and new sleep-state assignments for transistor stacks are proposed for total leakage minimization. We also present circuit re-organization schemes for total leakage reduction of dynamic circuits in sleep mode. Finally, we look into the effect of gate leakage on the MTCMOS circuit scheme and propose the use of sleep-state assignment in conjunction with MTCMOS to obtain increased total leakage savings.

II. GATE LEAKAGE ANALYSIS

Gate leakage current for an NMOS transistor of 0.1 μ m process shows an exponential dependence on the gate-to-source bias. At high gate bias, gate leakage current decreases with increasing drain-to-source bias. This can be attributed to the fact that a higher drain voltage results in a smaller electric field across the gate oxide at the drain end of the channel (lower VGD). At low gate bias, gate leakage was found to increase with increasing drain bias (due to the increase in reverse gate leakage with increasing drain bias, i.e., VGD). Thus, for a given gate-to-source bias, gate leakage is minimum when the gate-to-drain voltage is minimized. In addition, gate leakage current was found to be almost insensitive to the body-node voltage. The techniques presented in subsequent sections aim at minimizing the gate-to-source (VGS) and gate-to-drain (VGD) bias across a majority of devices, thereby obtaining a reduction in gate leakage and total leakage of the circuit.

III. STATIC CIRCUITS

Consider a three-high NMOS transistor stack (as found in the Nand3 cell shown in Fig. 1). The sub-threshold leakage

through the transistor stack is minimized when all of the devices in the stack are turned 'OFF', i.e., when a <000> pattern is applied. Since conventional leakage minimization techniques focus on sub-threshold leakage, the <000> pattern is believed to be the lowest leakage vector for a Nand3 cell. However, when such a pattern is applied, the output is high, and all of the PMOS devices experience high gate-to-drain and gate-to-source voltages. This results in a high field across the gate oxide causing gate leakage, which can be substantial due to the greater width of PMOS devices. To reduce gate leakage, it is necessary to

maintain the terminals of most of the devices at the same potential. This can be achieved by turning 'ON' all but the lowest NMOS transistor in the stack, i.e., by applying the input pattern <110>. Under such an input vector, only one PMOS device (P3) exhibits gate leakage. The gate leakage of the 'ON' NMOS transistors (N1, N2) is also negligible, since the internal nodes in the stack are charged almost to the supply rail (and hence the devices have a low VGS/VGD). The 'OFF' transistor (N3) at the bottom of the stack prevents subthreshold leakage from increasing tremendously. The total leakage for some of these vectors (fig-2) is clearly dominated by the gate leakage component. Even though sub-threshold leakage for the vector <110> is greater than sub-threshold leakage for the vector <000>, <110> is the minimum total leakage state for Nand3 cell. Thus, it is necessary to re-evaluate conventional leakage minimization schemes and input vector assignments to account for the effect of gate leakage. With gate leakage expected to increase more rapidly than sub-threshold leakage, we expect that turning 'ON' all but the lowest device in a transistor stack will be the lowest leakage state for a transistor stack in future technology generations.

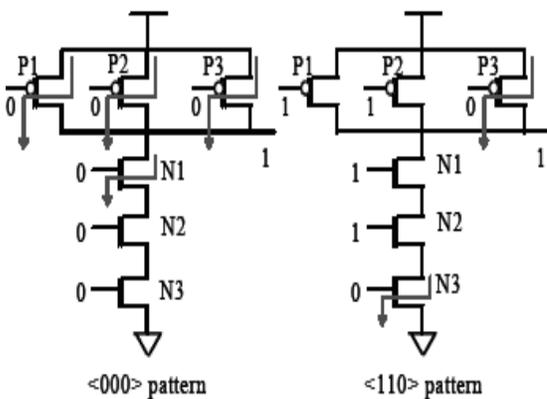


Fig.1 Input patterns for NAND3 cell

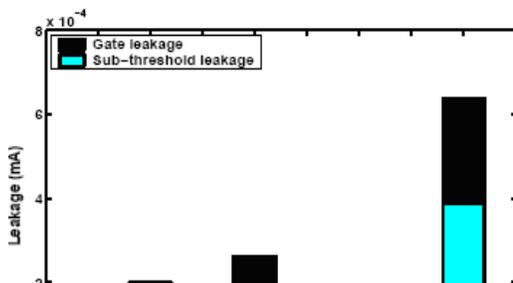


Fig-2 Total possible input vectors for NAND Cell refer to total leakage and its composition

IV. DYNAMIC CIRCUITS

This section focuses on sleep-state leakage minimization of dynamic circuits. Consider a typical 2-input dynamic AND cell as shown in Fig.3 During sleep state, the clock is held either in the precharge phase (low) or the evaluate phase (high). If the clock is held in the evaluate phase, the dynamic node will be discharged, and the output will be at logic high. Since, in a domino chain, the output of a dynamic cell drives other similar cells, it can be assumed that the inputs to the dynamic cell will also be at logic high. In such a state, all of the devices in the pulldown n-stack and the output pull-up transistor (i.e., devices on the evaluate path) will exhibit gate leakage. Since these devices are sized to reduce delay, it can result in significant gate leakage current. The sub-threshold leakage in this state is small, since it is primarily through the devices on the precharge path. On the other hand, when the clock is held in the precharge phase, the dynamic node is charged high, the output will be at logic low and the inputs can be assumed to be at logic low. In this case, the devices on the precharge path exhibit gate leakage, while the devices on the evaluate path contribute to the sub-threshold leakage. Though sub-threshold leakage of the NMOS pull-down tree is minimal due to stacking effect, sub-threshold leakage through the wide output pull-up transistor can be considerable. Thus, in either of the two states, the total leakage of the cell may be high, although due to different mechanisms. Conventional techniques claim that holding the clock in the evaluate phase is the lowest leakage sleep state, but this approach completely neglects gate leakage. Two proposed schemes are shown in Fig. 4. Both of these aim to minimize the total (sub-threshold plus gate) leakage current of the cell in sleep state. The output pulldown tree is modified to incorporate two small devices (N1, P1) that are controlled by the sleep-state control signal S. The precharge and evaluate clocks are separated in Scheme A. This will need additional circuitry for clock separation and can also result in clock skew problems. Scheme B uses a single clock, similar to the original configuration. In Scheme A, in sleep state, the precharge clock is held high, while the evaluate clock is held low. The inputs to the cell can also be assumed to be high. The activated conditional pull-up devices (P1, P2) therefore charge the dynamic node and the output to logic high. In this state, gate leakage of both the evaluate and the precharge paths are reduced (only the evaluate transistor exhibits reverse gate leakage

current) since most of the devices see an identical voltage at all of their terminals. The sub-threshold leakage of the output pull-up PMOS device is also reduced due to the 'OFF' device N1, resulting in significant savings in total leakage power. Since all of the additional devices are small, the delay degradation on the critical path is minimal. The additional devices can be desirably sized to obtain requisite precharge times and leakage savings. For Scheme B, the clock is held low in sleep state. The dynamic node and the output of the cell are high (similar to scheme A) reducing the gate and sub-threshold leakage of the output PMOS inverter. The savings in total leakage is slightly reduced, since the precharge transistor exhibits gate leakage in addition to an increase in the sub-threshold leakage through the evaluate tree. However, in this configuration, no additional devices are needed in the evaluate tree, minimizing the delay degradation. The percentage savings obtained in gate leakage power and total power, along with the area overhead and degradation in precharge and evaluate times, are listed in Table 1 for several commonly-used dynamic circuits. For instance, savings of over 73% in gate leakage and 13% in total leakage are obtained for the dynamic AND cell shown in Fig. 3,4,5 by using Scheme A with an area penalty of less than 7% and just over 1% degradation in delay. Fig 3,4 & 5 describe for Dynamic circuit reorganization for gate and total leakage minimization.

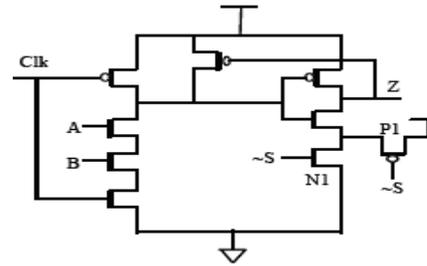


Fig-5 Scheme B, Single Precharge and Evaluate logic

V. MTCMOS CIRCUITS

The MTCMOS scheme has been proposed for reduction of sub-threshold leakage current in sleep state [5]. In this section, we investigate the effect of the MTCMOS configuration on gate and total leakage. The three configurations shown in Fig. 5 are considered. In the sleep state, the high VT footer and header devices are turned 'OFF' (thereby minimizing sub-threshold leakage current). This causes the virtual supply rails to be close to VDD or ground (if only footers or headers are used), or to be close to VDD/2 (if both headers and footers are used). The total leakage is the sum of the sub-threshold leakage of the sleep devices, the gate leakage of the sleep devices and the gate leakage of the input stage. The devices of the first stage may exhibit gate leakage depending on the input vector. For instance, if only footers are used, the virtual ground plane will be close to VDD. Thus, all of the devices in the logic circuit have their drain and source at nearly the supply rail. If an input vector of <0000> is applied, then all of the devices in the first stage will see a high VGS and VGD, and hence exhibit gate leakage. However, when an input vector <1111.> is applied, these devices will have identical voltage at all of their terminals, resulting in minimal gate leakage. This makes the leakage in sleep-state for the footer-only, configuration dependent upon the applied input vector.

A similar argument can be presented for the header only configuration. Here if Header only or footer only scheme is used, an appropriate input vector (00000... or 11111...) should be applied to obtain maximum savings in total leakage. This is validated in Fig. 6 which plots the gate leakage for an industry-standard decode circuit for each of the above schemes. In this case, the first 23 vectors are randomly generated, while vectors 24 and 25 are the <000...> and <111...> vectors. Here the total leakage for the header only configuration with the <000...> input vector is over 50% lower than the average leakage of remaining 24 vectors. Similarly, the application of the <111...> vector for the footer only configuration results in over 40% savings compared to the average leakage for the other 24 vectors. Further, the gate leakage of the sleep-state devices

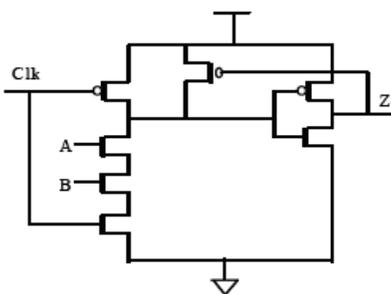


Fig-3 Typical two Input AND Gate

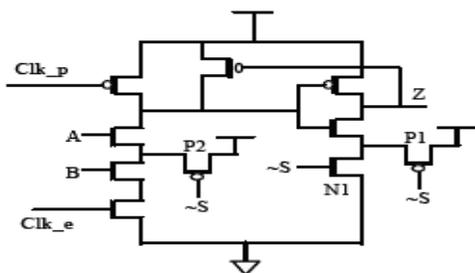


Fig-4 Scheme A, Separate Precharge and Evaluate logic

(headers and footers) can be significant since these devices experienced a high reverse VGS and VGD. leakage can be reduced by using both headers and footers. When both headers and footers are used, the virtual supply and ground rails float close to VDD/2. The gate-to-drain and gate-to-source bias across the sleep devices is reduced by about half, and hence their gate leakages are reduced. The gate leakage of the input stage is also reduced. Table 2 lists the ratio of the leakage currents for the MTMOS configurations of Fig. 6 compared to the leakage of the original circuit for an industry standard decode circuit in an advanced process.

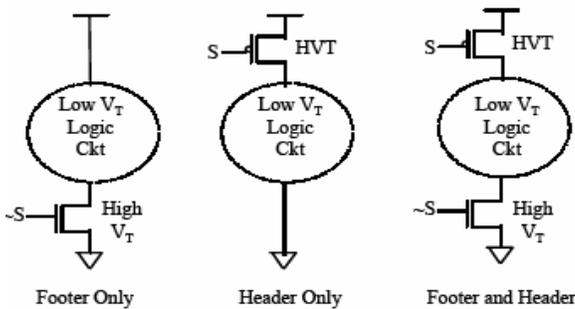


Fig-6 MTCMOS configuration (A, B, &C) evaluated

	Subthreshold savings	Gate leakage Savings	Total leakage savings
Footer Only	15.1	39.6	80.8
Header Only	111.0	28.4	61.2
Header & Footer	720.3	270.1	505.3

Table 2: Leakage currents ratio for MTCMOS configurations of Fig. 6 compared to original circuit

VI. CONCLUSIONS

In this paper we have shown the growing importance of gate leakage current and have clearly demonstrated the need to consider gate leakage in any leakage minimization scheme. We provided an analysis of gate leakage and presented optimal sleep-state assignments for transistor stacks in static circuits. We proposed new dynamic circuit

The savings in total leakage current using these schemes range from 2% to 38% with less than 7% increase in device area. Also try to evaluate the MTCMOS from a gate leakage perspective

and illustrated the need to use both headers and footers to obtain maximum leakage savings.

7. REFERENCES

[1] International Technology Roadmap for Semiconductors, 2001 Edition, <http://public.itrs.net/Files/2001ITRS/Home.html>
 [2] T. Inukai, et.al, "Boosted Gate MOS (BGMOS): Device/Circuit Cooperation Scheme to Achieve Leakage-Free Giga-Scale Integration," *Proc. CICC 2000*, pp.409-412
 [3] F.Hamzaoglu and M. Stan, "Circuit-Level Techniques to Control Gate Leakage for sub-100nm CMOS," *Proc. ISLPED*, pp. 60-63, Aug. 2002.
 [4] Y. Yeo, et.al, "Direct Tunneling Gate Leakage Current in Transistors with Ultrathin Silicon Nitride Gate Dielectric," *IEEE Electron Devices Letters*, vol.21, no.11, pp. 540-542, Nov. 2000.
 [5] S. Mutoh, et.al, "1-V Power Supply High-Speed Digital Circuit Technology with Multi-Threshold Voltage CMOS," *IEEE Journal of Solid State Circuits*, vol. 30, no. 8, pp. 847-854, Dec2003.

About author

Kanika Kaur (Associate Professor, KIIT, Gurgaon) received B.Sc (Electronics) Hons. Degree from Delhi University in 1997 and M.Sc (Electronics) Hons. Degree from Jamia Millia Islamia University in 1999. She received M.Tech degree from RTU in 2005 and presently pursuing Ph.D from the JJTU, Rajasthan in the field of "Low power VLSI design-subthreshold leakage reduction technique for CMOS". Published more than 20 research papers in national, international journal & conferences. She has also published a book titled "Digital System Design" by SciTech Publication in 2009. Editor of 05 Technical Proceedings of National & International Seminars. Convener of many National and International Symposium. Life member of IETE & ISTE. Awarded as best academic personality & HOD in 2007 and 2008 at NIEC, Delhi.

Table 1: Percentage savings and penalties for dynamic circuit reorganization schemes of fig.3, 4, & 5 compared to conventional dynamic circuit

Circuit	Scheme A					Scheme B				
	Gate leakage savings	Total Leakage saving	Evaluate Penalty	Precharge penalty	Area penalty	Gate leakage savings	Total Leakage saving	Evaluate Penalty	Precharge penalty	Area penalty
3 i/p And	73.50	13.0	1.20	27.90	6.00	51.3	14.0	-0.51	27.3	4.03

Design for Temperature-Controlled Solar Heated Chick Brooder

Okolie Paul Chukwulozoie¹, Okafor Emeka Christian^{2*}, Chinwuko Emmanuel Chuka², Ubani Nelson Obinna³, Okenwa Ugochukwu³

¹Correspondance Email: pc.okolie@unizik.edu.ng

Abstract— Brooding is the care of the chicks from a day old to six weeks. The current design consists primarily of the provision of modulated heat against the background of natural and artificial brooding system of brooding that is prone to fire hazard. This work therefore presents a design for a brooder that makes uses of solar as source of the required heat in the housing. The overall heat transfer coefficient of the door was determined as $3.846\text{W/m}^{20}\text{C}$ and transmission heat loss through door (Q_d) results to 30.15W . Then the heat transfer coefficient is 1.780 such that the total heat of 80.63W is loss through the walls. Based on the results, it was recommended that if there are several chicks, the temperature can usually be a little lower as the chicks keep each other warm and single chicks have fewer options to self-regulate; In many cases the heating can be moderated once the temperature was regulated down to 21 degrees Celsius.

Index Terms—Brooding, Energy, Heat, Insulation, Solar Collectors, Temperature.

1. INTRODUCTION

Nigeria receives about 4.851×10^{12} KWh of energy per day from the sun. This is equivalent to about 1.082 million tons of oil Equivalent (mtoe) per day, and is about 4 thousand times the current daily crude oil reduction, and about 13 thousand times that of natural gas daily production based on energy unit. According to Bala et al (2000), Nigeria is endowed with an annual Average daily sunshine of 6.25 hours, ranging between about 3.5 hours at the coastal areas and 9.0 hours at the far northern boundary. This huge energy resource from the sun is available for about 26% only of the day.

Solar chicken brooders of various sizes have been developed by national centre for energy research and development (NCERD), Nsukka. The chicken brooders use solar radiation as source of heat in place of electricity bulbs, kerosene lamps or stoves to provide heat for freshly hatched chickens. The technology eliminates the emission of product gases that are hazardous in health of both man and chicken (Iloeje, 1997). Some design however has provision for feeding and watering of chickens as well as collection and discharge of chicken droppings. Solar egg incubators are also available (Iloeje, 1997, Okeke, 2002, Oparaku, 2007).

Okonkwo and Akubuo (2001), utilized a water storage tank constructed with 2mm thick galvanized metal sheet measuring $3\text{m} \times 1\text{m} \times 2\text{m}$ insulated with wooden materials, 5cm thick to control the brother temperature. Studies

(Okonkwo *et al.*, 1993; Okonkwo and Agunwamba, 1997, Rockby *et al.*, 1983) showed that conventional energy consumption in poultry production is quite enormous and therefore expensive. Nonetheless, these studies did not include an explicit design consideration for temperature control and thorough housing material conductivity, a gap which the current study intends to fill.

2. MATERIAL AND METHODS

Selection of material of construction is dependent on the properties of the materials that give better performance. The design consisted of a flat plate collector mounted on a stand with variable angle of inclination, the thermal performance of a collector was determined using the first law of energy balance (Samtos et al, 2005). Comprehensive evaluation of environmental conditions around Awka-Anambra state Nigeria (the area where the research was carried out) informed the selection of design ambient temperature T_a to be 27°C ; Design brooder chamber temperature $T_b = 34^{\circ}$; Design brooder chamber relative humidity = 70%; Design ambient air relative humidity = 50%; The only heat gain considered is that due to metabolic process in the chicks. The heat (Q) is obtained from $= 3.5\text{Mass}^{0.75}$. The design model of (Duffie and Beckman, 1976; Duffie and Beckman 1991; Meinel and Marjorie 1976; Santos, Queiroz and Borges, 2005) as reported and validated by (Bello and Odey, 2009) were used in order to determine the heat transfer along the plate, an assumption

of temperature gradient being negligible is taken in the direction of flow, the rate of useful energy gain (W) was then computed.

3. RELEVANT MATHEMATICAL FORMULATIONS

The energy loss at a given temperature depends on the overall energy loss coefficient, U_c and the area of the collector through which heat is lost. The energy loss coefficient could be reduced by second glazing of the absorber plate or by surrounding the collector plate by a vacuum.

$$q_u = I_c A_c I \alpha_s \tau_s - U_c A_c (T_c - T_a) \tag{1}$$

Where A_c = Area of the collector absorber plate (m^2); I_c = Global insulation on plane of the collector (W/m^2); τ_s = Net solar transmittivity of glazing; α_s = Solar absorptivity of collector plate; U_c = Overall heat loss coefficient ($W/m^2 \cdot C$); T_c = average collector plate surface temperature ($^{\circ}C$); T_a = Ambient air temperature ($^{\circ}C$). Thermal efficiency η_c is given by

$$\eta_c = \frac{q_u \eta_0}{A_c I_c} - U_c \frac{(T_c - T_a)}{I_c} \tag{2}$$

Where; η_0 = the optical efficiency of the collector, equivalent to the product; $\alpha_s \tau_s$ in equation (1). For a simple flat-plate configuration, the optical efficiency; η_0 = (cover transmittivity) x (solar absorptivity); for which the value of τ_s ranges from 0.75 - 0.95 and α_s from 0.90 - 0.95. For concentrating collectors;

$$\eta_0 = \rho_s \tau_s \bar{b} \tag{3}$$

Where; ρ_s = Solar Reflectivity of the concentrator reflector surface; \bar{b} = intercept factor.

Collector Plate Temperature, T_c is eliminated from the efficiency relation by using by using collector heat removal efficiency factor, F_R . This because of the difficulties encountered in trying to obtain the collector temperature.

$$F_R = \frac{I_c \tau_s \alpha_s - U_c (T_c - T_a)}{I_c \tau_s \alpha_s - U_c (T_{fi} - T_a)} \tag{5}$$

$$Q_u = F_R A_c [I_c \tau_s \alpha_s - U_c (T_{fi} - T_a)] \tag{6}$$

In order to avoid admission of surplus radiation into the housing, the design excluded window on the east and west walls. However, two windows are required for the proper ventilation on the south and north walls.

Considerations for transmission heat loss through the floor and wall

The floor comprises sheet of plywood with thermal conductivity of $0.118W/m^{\circ}C$ and thickness $0.005m$, hard wood of thickness $0.04m$, and insulation material of wood fiber (soft wood) of thickness $0.06m$ and thermal conductivity of $0.043w/m^{\circ}C$

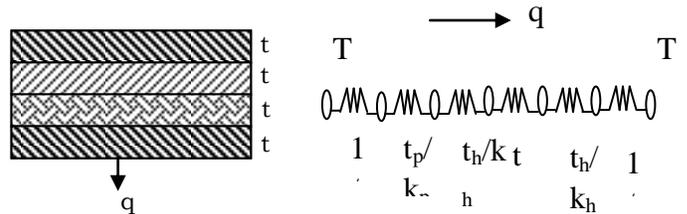


Fig. 1: heat transfer section of the floor and network

The overall heat transfer coefficient, U_f is therefore obtained from (7);

$$1/U_f = 1/h_b + \frac{t_p}{K_p} + \frac{t_h}{K_h} + \frac{t_s}{K_s} + \frac{t_h}{K_h} + \frac{1}{h_a} \tag{7}$$

Heat loss from the floor, Q_f is

$$Q_f = A_f U_f (\Delta T)_f \tag{8}$$

The walls comprise hardwood and insulation wood fiber as in the floor.

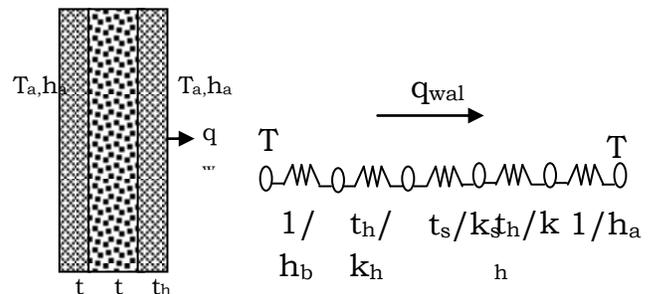


Fig 2. Heat transfer section of the wall and network

Considerations for transmission heat lost through the roof

The roof is made corrugated iron sheet inclined at 16.22° . According to (Oguike 1991), with the assumptions that the thermal resistance of the roofing material is negligible considering its thickness which is very thin, and its high thermal conductivity; There is a ceiling that shields the room from the roof; therefore heat loss by transmission is only through the roof.

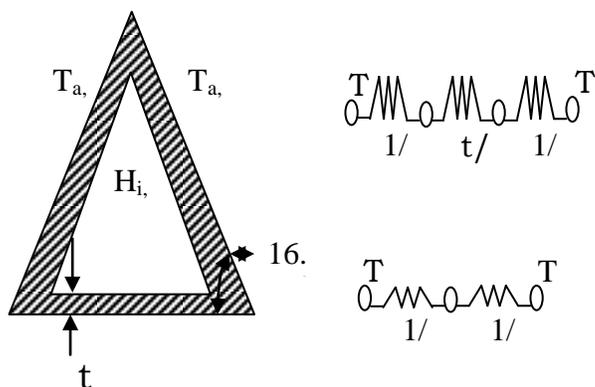


Fig 3. Heat transfer section of the roof and network

The heat transfer through the roof is given by

$$Q_r = UA_{cel}(T_b - T_a) \tag{9}$$

$$Q_r = U_{cel}A_{cel}(T_b - T_i) \tag{10}$$

$$Q_r = U_r A_r (T_i - T_a) \tag{11}$$

$$U = \frac{1}{U_{cel} + (A_{cel}/U_r A_r)} \tag{12}$$

Where U_{cel} = Heat transfer coefficient from brooder to attic; U_r = Heat transfer coefficient attic to ambient; U = Overall heat transfer coefficient through the roof; A_{cel} = Area of the ceiling; A_r = Area of the roof; T_b = Boarder house temperature; T_a = Ambient temperature ; T_i = Attic air temperature. According to (Ashrae, 1981), air exchange per hour is used for residential building because of the average sized windows. This design specified smaller windows such that air method becomes insufficient. Therefore, crack length method is used and rate of air flow through closed windows is given as:

$$V = KP^n \tag{13}$$

Where K = proportionality constant; P = pressure difference across window; n = exponent flow which is between 0.5 to 1.0

4. RESULTS AND DISCUSSION

The overall heat transfer coefficient of the door was determined as $3.846W/m^2C$ and transmission heat loss through door (Q_d) results to $30.15W$. Then the heat transfer coefficient is obtained from (7) as 1.780 such that the total heat loss through the walls is $Q_{wall} = 45.30 + 35.33 = 80.63W$. Considering ventilation requirement for poultry, using small wooden windows, and based on the optimum air exchange rate of poultry and velocity of air flow of $0.85m^3/hr$ bird and $0.4m/s$ respectively, the air flow rate for 100 chicks in m^3/s is found to be $0.0236m^3/s$. Thus each window will be of area = $0.0295m^2$. Assuming a square window thickness, $0.02m$, Length of window = $0.172m$, Width of window = $0.172m$. A single door is required and located on either south or north wall. The dimension of the door is thickness = $0.02m$, height $1.4m$, width $0.7m$.

Table 1: design specifications for solar heated chick brooder

Medium of Heat Loss	Heat transfer coefficient (U) W/m^2C	Area A m^2	Total heat loss Q (W)
Windows	3.846	0.1183	3.64
door	3.846	0.98	30.15
Floor	0.5486	5.02	19.28
Walls	0.5617	8.96	80.63
roof	2.6682	5.02	227.46

The average mass of the bird was calculated to be 1.045 and U_t gave $1.956W/m^2 C$ and thickness is $0.0324m$, thus the optimum insulation thickness is specified as $3.5cm$ or $35mm$. The air flow rate per hour associated with weight of a bird at each age per bird and the velocity is recorded and presented in table 2, specifying the weight of bird at each age and ventilation requirement for brooder.

Table 2. Recorded weight of a bird at each age of brooding

Age of fous (week)	ventilation ($m^3/Bird/hour$)		air moveme nt	averag e mass (g)
	Min	Max		

			velocity (m/s)	
Week 1	0.162	1.296	0.2	37.5
Week 2	0.324	1.296	0.3	92.5
Week 3	0.456	1.296	0.3	189

Type of material	Material grading	Specific Infiltration (m ³ /s-m ² frame)		
Wood frame small	MTRGA	0.01016		
Wood frame, weather stripped	MTRGB	0.00381		
Wood frame large	MTRGC	0.00762		
Metal frame small	MTRGD	0.00762		
Metal frame, weather stripped	MTRGE	0.00381		
Metal frame large	MTRGF	0.00508		
Week 4	0.612	3.888	0.4	328
Week 5	0.695	4.248	0.4	556
Week 6	0.817	6.48	0.4	688

In a similar concern, the variation of surface conductance with air velocity is represented in figure 4. Indicating a near linear property.

Table 3. Infiltration through windows (m³/s-m²frame)

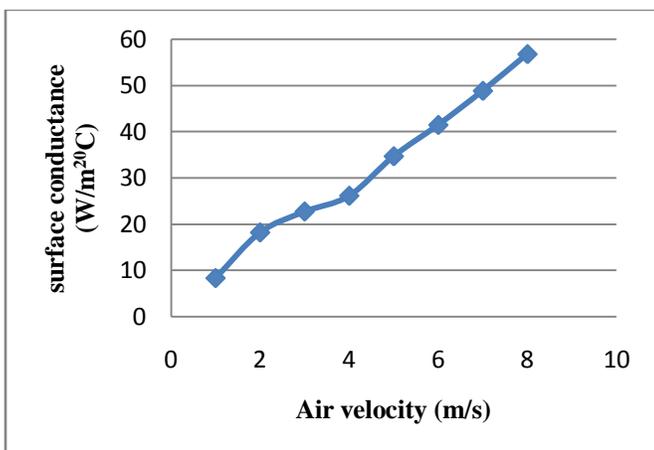


Fig. 4: Variation of surface conductance with Air velocity

For the brooder air velocity of 0.4m/s, by interpolation, $h_b = 10.05 \text{ W/m}^2\text{°C}$. Using the world recognized outside standard velocity of 5m/s, then $h_a = 28.143 \text{ W/m}^2\text{°C}$. Thermal conductivity for hard wood, $k = 0.16\text{W/m}^2\text{°C}$. The overall heat transfer coefficient of the network is $U_w = 3.846\text{W/m}^2\text{°C}$. According to (Doserient) provided the room is painted white or light colour, the following correction factors can be applied for temperature; East wall 2; West wall 2; North wall 1; South wall 1. Then, the total heat transmitted through the window is such that total heat transmitted through the window as shown in fig 6 is 3.64W, while total rate of infiltration through the windows as shown in table 3 is $1.202 \times 10^{-3}\text{m}^3/\text{s}$. For the four wooden small frame-windows. Rate of infiltration through the door becomes $2.529 \times 10^{-5}\text{m}^3/\text{s}$, giving a total infiltration of $1.227 \times 10^{-3}\text{m}^3/\text{s}$.

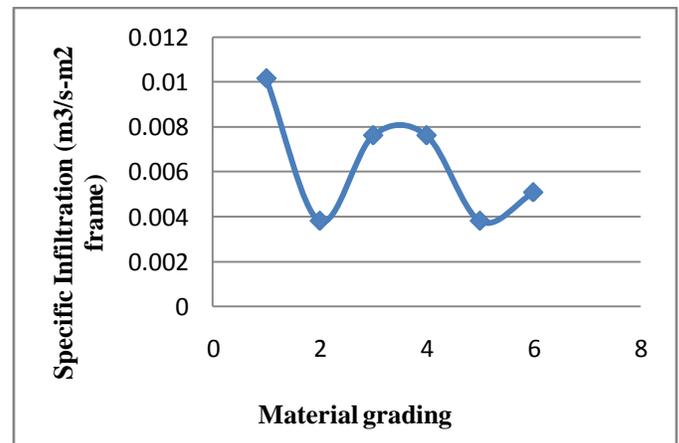


Fig. 5: Recorded Infiltration through windows

This recommended average rate of infiltration value of crack is because infiltration also depends on rate of door opening and closing. Air enters through the direction of the wind and leaves through the leeward opening by the side with the highest opening; the total rate of infiltration through the windows therefore is $1.202 \times 10^{-3}\text{m}^3/\text{s}$ and for the four wooden small frame-windows, the rate of infiltration through the door is $2.529 \times 10^{-5}\text{m}^3/\text{s}$ hence total infiltration was computed as $1.227 \times 10^{-3}\text{m}^3/\text{s}$.

Similarly, the heat loss by infiltration considered as the sum of sensible heat load and latent heat load for which the rate of infiltration in the brooder gave $1.227 \times 10^{-3}\text{m}^3/\text{s}$ and density of air in the brooder house at atmospheric

pressured of $1.013 \times 10^5 \text{N/m}^2$. Assuming an ideal gas for air, then total heat loss from the brooder resulted 285.30W and heat load of the brooder gave 113.23W

From the heat transfer network of the ceiling and attic, the ceiling thickness of 0.015m and ceiling thermal conductivity of $3.32 \text{W/m}^\circ\text{C}$ were employed for the computation of total heat loss through the roof and transmission that equals 227.46W, however, the door is of the same thickness and material as the window, and exposed to the same brooder condition and outside ambient condition. Thus the overall heat transfer coefficient of the door is $3.846 \text{W/m}^2\text{C}$

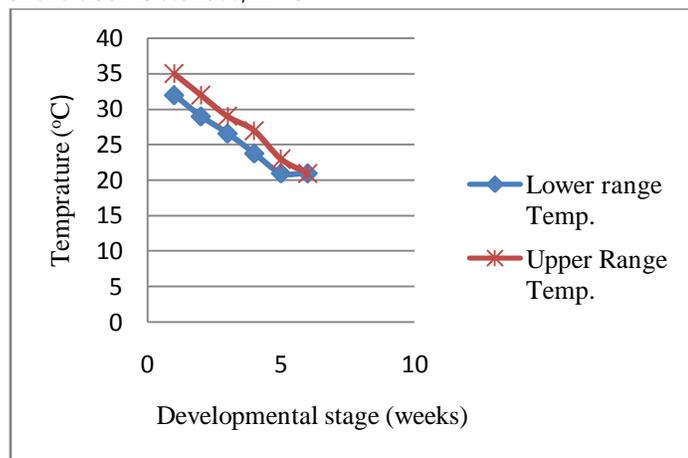


Fig 6. The upper and lower range brooder temperature variably decreased as the chick develops

In general, parent hatched chicks are likely to be more comfortable at lower brooder temperatures due to the hen occasionally leaving the nest. However, the prevailing weather conditions should be taken into account. Thus the design implementation maintained a warm brooder up to the correct temperature before adding the chicks. A non-breakable thermometer at "chick level" was used to obtain the temperature value in table 4, which can serve as a guide temperature to monitor the temperature; the results were compared with (Oguike, 1991) and found reliable.

Table 4. Recorded results of chick brooder temperature

Age of Poultry (feathered) Chick	1st Wee k	2nd Wee k	3rd Wee k	4th Wee k	5th Wee k	6th Wee k +
Temperatur	32 -	29 -	26.6 -	23.8 -	21 -	21

e °C	35	32.	29	27	23	
------	----	-----	----	----	----	--

Collector Specification

The average daily irradiation of the location over the year is given this value is used to specify the area of the flat plate collector under the calculated heat load. Form collector area required. The total heat required = heat load = 120.18W. Heat from collector and collector efficiency, assuming 40% for a simple cover plate; Heat form collector therefore is specified as 1689.82W, such that considering a square plate with length of 0.27m; Width of 0.27m; Area of 0.0729m^2 and absorber plate made of steel plate with thickness 0.005m. The plate is coated with a flat black paint with absorptivity and emissivity of 0.95 and 0.97 respectively. A single green house grade fiber glass of transitivity 0.80, and emissivity of 0.90 serves as the top cover. The insulation is made of saw dust material because its availability. The enclosure is made from corrugated iron sheet material, the selection of materials of the both collector is based on the economic important and the availability of the materials.

5. CONCLUSIONS

There is abundance of sunshine in Africa as a constant and in Nigeria as country. This solar energy can be harnessed for human use. The study gave an overall heat transfer coefficient, U_f of $0.5486 \text{W/m}^2\text{C}$ and Heat loss from the floor, $Q_f = 19.28 \text{W}$. The idea of this work is to make it very cheap for the local farmers to use manure to reduce drastically, the cost of energy in agriculture. In this paper effort has been made to use local material so as to reduce the cost that will be involved in the construction to the tune of about NGN 150,000.00 depending on the prevailing market condition.

Also the design was made taking into consideration the local condition of the rural areas and enough safety factor applied to ensure reliability of the brooder chamber. The temperature control ensures that maintenance of temperature between 34°C to 24°C can be achieved. The brooder can further be optimized for temperature-controlled enclosures for hatching and raising baby birds. It can be extended for used by agricultural breeders of poultry, commercial purveyors of domestic birds such as parakeets and canaries, and wildlife rescue. It can also be employed by rehabilitation and repopulation scientists for saving endangered species, and for individuals who make

small-scale purchases of baby chickens and ducklings or need to nurse a sick or injured bird. Finally, if there are several chicks, the temperature can usually be a little lower as the chicks keep each other warm. Single chicks have fewer options to self-regulate. In many cases the heating can be removed once the temperature was regulated down to 21 degrees Celsius, however, if outside temperatures are much cooler, then the lamp needs to stay on and temperatures have to be slowly reduced to be close to that of the temperature they are going to be maintained in.

REFERENCES

- [1] Ashrae (1981) Hand Book of Fundamentals, "American Society of Heating Refrigerating and Air conditioning Engineers, inc., New York.
- [2] Bala, E.J., Ojosu, J.O. and Umar, I. H. (2000). Government Policies and Programmes on the Development of Solar-PV Sub-sector in Nigeria. Nigerian Journal of Renewable Energy, Vol. 8, No. 1&2, pp. 1-6
- [3] Duffie, J. A and Beckman, W.A.,(1976). Solar collectors in clear sky conditions *J. agric res. Vol. 28*, pp 321-328.
- [4] Duffie, J. A., Beckman, W. A. (1991). Solar Engineering of Thermal Process, 2nd ed., New York, John Wiley.
- [5] Ileoje, O.C. (1997) Potentials for Renewable Energy Application in Nigeria, Energy Commission of Nigeria, 5 – 16.
- [6] Meinel A.B. and Marjorie P. Meinel, (1976). Applied Solar Energy- an introduction. Addison Wesley publishing company inc. NY.
- [7] Oguike, U.A. (1991). Design of solar chick brooder," An unpublished B. Eng. Project; Department of Mechanical Engineering, university of Nigeria Nsukka, July.
- [8] Okeke C.E. (2002). Activities at a Glance, The Presidency Energy Commission of Nigeria, National Centre for Energy Research and Development, University of Nigeria, Nsukka, 3 – 5.
- [9] Okonkwo W.I and Akubuo C.O.(2001). Thermal Analysis and evaluation of heat requirement of a passive solar energy poultry chick Brooder. *Proceedings of NIAE Vol.23*, pp 373.
- [10] Okonkwo, W.I and J.C. Agunwamba, (1997). Socio Economic Impact of Solar Brooding System, Proceedings. International Conf. On Power Systems Operation and Planning, Abidjan.
- [11] Okonkwo, W.I., O.C. Ileoje, E.A. Echiegu and C.O. Akubuo, (1993). Performance evaluation of a Medium Scale Passive Solar Heated Poultry Chick Brooder, Nig. J. Solar Energy.
- [12] Oparaku, O.U. (2007). Activities at a Glance, Energy Commission of Nigeria, National Centre for Energy Research and Development University of Nigeria, Nsukka, 7 – 20.
- [13] Rockby, T.R, J.M. Redford and D.J. Pitts, (1983). Solar Heating for Commercial House: A Further Evaluation Trans. ASAE.
- [14] Santos B.M. Queiroz M. R. and Borges, (2005). A solar collector design procedure for crop drying. *Brazilian journal of chemical engineering. Vol. 22. No. 02* pp. 277-284, April-June, 2005.
- [15] Bello, Segun R. and Odey, Simon Ogebeche (2009) Development of Hot Water Solar Oven for Low Temperature Thermal Processes. Leonardo Electronic Journal of Practices and Technologies. p. 73-84.

Critically Analyzing the Behavior of Fuzzy Flip Flops and Searching for Better Solutions

SABYASACHI CHAKRABORTY

Abstract— Sequential circuits are the important components of a computer system as they process binary data and buffers binary output for the next period.. Sequential circuits are the interconnection of logic gates and Flip Flops and the buffering of binary output for the next period is carried out by the Flip Flops by their inherent feedback mechanism. Sequential circuits basically process binary data sequentially with an automated method of state transition which is solely driven by the Flip Flops. Among the different Flip Flops J-K Flip-Flop is the major one and the most elementary and general unit of the sequential digital circuits. In this paper the J-K Flip Flops driven by Fuzzy rules are viewed as an improvised form of the Flip Flops based on the Boolean rules and the various operative aspects of them are examined. Firstly the detailed comparison between Fuzzy logic and Boolean axiomatic properties are carried out, then the Fuzzy logic driven Flip Flops with non associative properties are studied and the input-output relationships as well as the expressions of outputs in the next period are derived considering different combinations of the outputs in current period and inputs in the next period. At last the rules of output and different input conditions are analyzed and derived. The later part of the paper deals with the unstable situation of Fuzzy J-K Flip Flops over the different states and provides with some measure to avoid the same. Thus this paper deals with the formulation of fuzzy operational rules for the J-K Flip Flop on the one hand and attempts to find a concrete solution to the problem of unstable states of J-K fuzzy Flip Flops on the other.

Index Terms— Fuzzy J-K Flip Flop, Sequential Circuit, Triangular Norm, Triangular Co-norm, SET type Flip Flop, RESET type Flip Flop, Unstable Situation, Boolean.

◆

1 INTRODUCTION

In this paper the elementary results of analyzing Fuzzy sequential circuit with J-K Flip-Flop is presented. Firstly the classic Fuzzy operations and the non-associative operations are examined, secondly the Boolean axiomatic properties of the J-K Flip Flops are examined in comparison with Fuzzy and later the Fuzzy J-K Flip Flops with non-associative properties are examined. It has been studied that the triangular norm(t-norm) and triangular co-norm(t-co norm) or s-norm yield different expressions for SET and RESET type respectively are proved to be equivalent and thus yield a common single expression for both the norms. As the finding of the author, the rules are derived for the output at t+1 th period at different combinations of inputs and different SET-RESET conditions of the current period. In another study it has been found that the unstable situations arise with toggled outputs in the successive states of Fuzzy J-K Flip Flop. This problem is revealed by one of the reviewed papers which has been attempted to be resolved here by a circuit mix that aims to bring steady output levels across the states with a given fixed inputs.

2 ANALYSIS STAGE

2.1 Literature Review

(1) In the paper "Non-Associative Fuzzy Flip-Flop with Dual SET-RESET Feature" by Rita Lovassy; Institute of Microelectronics and Technology, Budapest Tech, Hungary. and Laszlo T. Koczy; Institute of Information Technology,

Mechanical and Electrical Engineering, Szechenyi Istvan University, Gyor, Hungary; Department of Telecommunication and Media Informatics, Budapest University of Technology and Economics, Hungary presented in the SISY 2006 - 4th Serbian-Hungarian Joint Symposium on Intelligent Systems, at first the classic Fuzzy operations and then the non associative operations are analyzed with some relevant example related to probability. In the later part the comparison between Fuzzy and Boolean flip-flops are carried out and finally the Fuzzy properties of non associative Fuzzy Flip Flops are analyzed and some important results are found. In the reviewed paper, the analysis starts with the minimal disjunctive(Sum of product) form as: $Q(t+1) = J \cdot Q(t)^c + K^c \cdot Q(t)$. Likewise the equivalent max term expression or minimized conjunctive form (product of sum) as: $Q(t+1) = (J + Q)(K^c + Q(t)^c)$

In the paper it is stated that per Boolean algebra these two are equivalent but in Fuzzy logic there is no such operation triplet for which these two are equivalent. There is no such mechanism also to prefer any one of them over the other. Thus in the reviewed paper, the concept of Hirota and Ozawa to use these two dual definitions of Fuzzy J-K Flip Flop is used. The first equation of $Q(t+1)$ is referred by them as the output equation of the "RESET" type Fuzzy Flip Flop and the later is referred as the output equation of the "SET" type Fuzzy Flip Flop. It is clear that the maxterm of the first equation is the dual of the first equation and physically can be achieved by complementing J and K inputs with inverters.

In the reviewed paper, the attempt is made to find out a single symmetric solution combining SET and RESET type equations with non associative properties. The non associative Fuzzy Flip Flop is discussed with a reference to Fodor and Koczy. The formulation is as follows:

$$P1 : F(0,0,Q) = Q,$$

$$P2 : F(0,1,Q) = 0,$$

$$P3: F(1,0,Q) = 1,$$

P4: $F(1,1,Q)=n(Q)$ where n is the strong negation of Q which is consistent to the Boolean concept of toggling of t th period's output in t+1 th period where both J and K are 1.

P5: $F(e, e,Q)=e$ where $e =n(e)$ is any equilibrium and n is the strong negation.

Combining P2, P3 and P5 we get,

$$P(6)= F(D,n(D),Q) =D \text{ where } D \in [0,1].$$

These all formulations are consistent with the Boolean logic identities. Now according to the Lucasiewicz norms the output of the RESET type flip-flop :

$$Q_R (t+1) = \min [T(J,1-Q) + T(1-K,Q),1].$$

Or we can write,

$$Q_R (t+1) = T(J,1-Q) + T(1-K,Q) \text{ -----}1.$$

Similarly for SET type Flip Flop,

$$Q_S (t+1) = \max[S(J,Q)+S((1-K),(1-Q))-1,0]$$

Or we can write,

$$Q_S (t+1) = S(J,Q)+S((1-K),(1-Q))-1 \text{-----}2.$$

We can clearly see that the output equation of RESET type Flip Flop is represented by Fuzzy T-norm and SET type by Fuzzy S-norm or T-co norm.. Now from 1 and 2,

$$Q_R(t+1)=\{\min(J,1-Q)+\max(J-Q,0)\}/2+\{\min(1-K,Q)+\max(Q-K,0)\}/2 \text{ -----}3.$$

This is achieved by using the mid-point based numerical estimation based on

$p= [\{\max(P_1+P_2-1),0\},\{\min(P_1,P_2)\}]$. As $T(J,1-Q)$ and $T(1-K,Q)$ are Fuzzy T-norms or or intersection, that's why this formulation is used. Similarly, using the concept of UNION of subjective probabilities $p=[\{\max(P_1,P_2),\{\min(P_1+P_2),1\}\}]$ and by using the midpoint based numerical estimation,

$$Q_S (t+1)= [\{\max(J, Q)+\min(J+Q,1)\}/2 + \{\max(1-K,1-Q)+\min((1-K+1-Q),0)\}/2] -1$$

Or

$$Q_S (t+1)= [\{\max(J, Q)+ \max(1-K,1-Q)\}-1]/2 +$$

$$[\{\min(J+Q,1)+\min((2-K-Q),1)\}-1]/2 \text{ -----}4$$

Now in the reviewed paper it has been proved that $Q_R (t+1) =Q_S (t+1)$ and they give a single expression for the output of Fuzzy flip-flop. Finally it has been implicitly proved that the SET and RESET equations are equivalent in Fuzzy flip-flop and thus there is no distinction between SET type and RESET type. A single expression of Fuzzy flip-flop output for different combinations of J,K and Q(t) is achieved.

It has been proved that

$$\begin{aligned} &\min(J,1-Q)+\max(J-Q,0) + \min(1-K,Q)+\max(Q-K,0) \\ &= \max(J, Q)+ \max(1-K,1-Q) \\ &+ \min(J+Q,1)+\min((2-K-Q),1)-2 \end{aligned}$$

Now there are $3! \cdot 2^3 =48$ combinations of J, K, Q(t), $J^c K^c Q(t)^c$. These 48 cases are not all essentially different. Any variable and its negated are symmetrical to the equilibrium $e=0.5$. Consequently, for describing a case it is sufficient to tell which one of the ponated or negated version of each of the three variables is less, greater or equal than e. The 8 main cases to be considered are as follows: $J^c K^c Q(t)^c$, $J^c K^c Q(t)$, $J^c K Q(t)^c$, $J^c K Q(t)$, $J K^c Q(t)^c$, $J K^c Q(t)$, $J K Q(t)^c$, $J K Q(t)$. For each combination there are $3! =6$ sub cases depending on the sequence of these three. Thus these are total 48 sub cases where some sub cases are giving identical results for Q_R and

• *Sabyasachi Chakraborty is currently working as Sr.Lecturer(IT/CS) and Head of the Department (IT/CS) in Icfai University-Sikkim-India and currently pursuing Ph.D in Application of Fuzzy Logic in Computer Science and Engineering from West Bengal University of Technology-India Email: scit75@gmail.com, Ph:919635107041. The author has performed and guided over 15 software engineering and research projects.*

Q_s . It has been observed that these are only 13 out of 48 sub cases are essentially different and those are listed in the table :

The first column of the table contains the serial number of the essentially different sub-case; the second column describes the inequality conditions applying for the given essential sub-case, while the third column gives the identical value of 2Q_R and 2Q_s in the given sub-case. Now it is clear that for all these essentially different sub cases, Q_s and Q_R gives single expression (as shown in the third column) and so it has been concluded that Q_s and Q_R are identical. Now some concrete rule base and conditional formulation should be there to explain these cases in a well formatted way.

(2) In the paper "New Components for building Fuzzy logic circuits" the authors Ben Choi and Kunal Tipnis of Louisiana Technological University has pointed out the case of several unstable conditions arise for the SET type and RESET type J-K Fuzzy Flip Flops. Some examples are provided as:

Initial Q	J	K	Q(t)	Q(t+1)	Q(t+2)
0	4	4	4	2	4
2	4	4	4	2	4
4	4	4	2	4	2
6	4	4	2	4	2

Here 6 Volt represents logic 1 and 0 volt represent logic 0 and 2 Volt and 4 Volt represents logic level .334 and .667 respectively.

While the initial stored value of Q is 0 volt and when given 4 volts of inputs for J and K , the resulting q will be continuously toggling between 2V and 4V. Similar types of unstable condition will appears when the initial stored value is 2 Volts and given 4 Volts of inputs for J and K. Other unstable conditions are also stated to take place when J=K=6 , J=4 k=6 and J=6 K=4.

It has been pointed out that neither the SET type nor the RESET type can solve the problem of unstable state. In this paper they used Hirota's technique of combining the SET type and RESET type equation is used.:

$$Q(t+1) = \{ J \vee (1-K) \} \wedge \{ J \vee Q(t) \} \vee \{ (1-K) \vee (1-Q(t)) \}$$

This combined equation minimizes the unstable states but can't completely solve it as there are still some unstable states

observed. As a suggestive measure the new Fuzzy D type Flip Flop has been referred as in D type Fuzzy Flip Flop J and K inputs can be complements to each other only. Other

Case	Conditions	2 Q _R (t+1)= 2 Q _s (t+1)
1	J, K <= Q, ¬Q	J- K +2Q
2	J, ¬K <= Q, ¬Q K, ¬J <= Q, ¬Q	1+J-K
3	¬J, ¬K <= Q, ¬Q	2+J-K-2Q
4	J, Q <= K, ¬K	J+Q
5	J, ¬Q <= K, ¬K	1+J-2K +Q
6	¬J, Q <= K, ¬K	1+J-Q
7	¬J, C <= K, ¬K	2+J-K-Q
8	K, Q <= J, ¬J	2J-K+Q
9	K, ¬Q <= J, ¬J	1-K+Q
10	¬K, Q <= J, ¬J	1+2J-K-Q
11	¬K, ¬Q <= J, ¬J	2-K-Q
12	Q <= J, K, ¬J, ¬K	2J
13	¬Q <= J, K, ¬J, ¬K	2-2K

combinations are ignored. This will eliminate the number of unstable states to some extent.

2.2 Finding of the Author

The paper by Lovassy and Koczy follows the process mentioned in the paper by Horota and ozawa in their works on the related topics. Basically it is shown that SET Type and RESET Type Flip Flop equations are equivalent and giving 13 distinguished subcases with combinations of two current period's inputs and one previous period's output. In the first reviewed paper the triangular norm and co-norm forms of Fuzzy Flip Flop operations are proved to be equivalent. However a SET of rules are further required to define the input output relations and the formulation of the outputs. Here a SET of rules and formulations are derived those are consistent for both "SET type" and "RESET type" Fuzzy Flip Flops. In the send paper reviewed points out the unstable situations and the new D type Fuzzy Flip Flop is referred as suggestive measure. However considering the power and scope of J-K Flip Flop in designing Fuzzy sequential circuit, it should be used as a prime component and D Fuzzy Flip Flop can hardly substitute it in every respect . The formulations are made to make the input-output relationship completely conditional and can clarify the unstable states as well. The formulations are as follows:

Case 1 :If J=K

Subcase-i: J=0,K=0

Q(t+1)=Q(t) – No Change , regardless of Q(t).

Subcase-ii: $J < .5$ and $K < .5$

If $Q(t) \geq .5$ then

If $Q(t) < J$ then

$$Q(t+1) = Q(t) \quad (\text{No change})$$

$$Q(t+1) = J$$

Otherwise

Otherwise

$$Q(t+1) = \neg Q(t) \quad (\text{Toggle})$$

If $Q(t) \geq J$ and $Q(t) \leq \neg J$ then

Subcase ii: $J < K$

$$Q(t+1) = Q(t) \quad (\text{No Change})$$

$$Q(t+1) = J.$$

Otherwise

$$Q(t+1) = \neg J.$$

This will view the system from logical point of view and the dynamic changes of output across the states are also clarified. The unstable states are all conditional and thus can be avoided instead of cutting down its scope (by switching to D Fuzzy Flip Flop).

Subcase-iii. $J = .5, K = .5$

$$Q(t+1) = .5 \text{ regardless of } Q(t).$$

These results are achieved both for $Q_R = (J \wedge \neg Q) \vee (\neg K \wedge Q)$. And $Q_S = (J \vee Q) \wedge (\neg K \vee \neg Q)$ and by using Fuzzy MAX-MIN logic. Thus these rules are consistent for SET and RESET type Fuzzy Flip Flops. Now only two cases are considered: where $J = K$ and where $J \neq K$. These formulations are derived empirically as well as by following the output rules of Flip Flops supported by Fuzzy SET theory. Here cases and subcases define accurately all the conditions represented by the combinations of inputs and output revealed in the paper by Lovassy and Koczy.

Subcase-iv. $J > .5, K > .5$

If $Q(t) < \neg J$ then

$$Q(t+1) = J$$

Otherwise

If $Q(t) > J$ then

$$Q(t+1) = \neg J$$

Otherwise

$$Q(t+1) = \neg Q(t) \quad (\text{Toggle})$$

- 2) In the paper by Ben Choi and Kunal Tipnis it has been pointed out the existence of unstable conditions in both SET type and RESET type flipflops. It has been examined that the technique applied by Hirota can reduce the no of unstable states but can't solve the problem. The suggested solution as derived in the research work is to feed the complemented Fuzzy input to J and K input lines from the 2nd state onwards whereas in the initial state the unmodified value of the original output should be taken. This can be done by using a counter. The count enable line of the said counter should be connected with the clock transition detector directly and the J-K Flip Flops of the counter will be provided with 0 or 1 signal (or 0V and 6V) only. Here the clock transition detector and counter follows Boolean logic but doesn't contradict with Fuzzy in the sense that they can be seen as working with terminal Fuzzy values (0 and 1 or 0V and 6V). Now as the first clock transition that will yield the output in t(1) period should supply unmodified inputs but from the second transition onwards the complemented outputs are to be supplied as inputs. A Fuzzy OR gate should be connected to all the output lines of the J-K Flip Flops of the counter except the first one (that represents right most bit). The output line of the Fuzzy OR gate is to be fed to the input lines of two Fuzzy XOR gates whose other

Subcase-v. $J = 1, K = 1$

$$Q(t+1) = \neg Q(t) \quad (\text{Toggle}).$$

Case -2: $J \neq K$ and $J = \neg K$

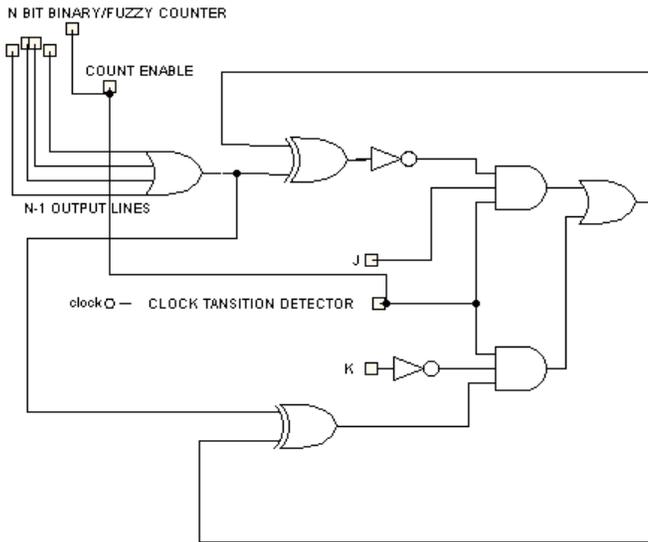
Subcase i: $J > K$

If $Q(t) \in (0, 1)$ then

$$Q(t+1) = J$$

Otherwise

inputs are taken from the output(Q(t)) of the previous period. This system will complement the output of the previous state except in case of initial state. By doing so the unstable conditions can be avoided and the Flip Flop will produce the same output volt over the different states. This is empirically proved for the combinations in the table of the concerned reviewed paper. The suggested circuit is as follows:



Here the n-1 output lines(except the first that represent rightmost bit) are inputted to a Fuzzy OR(equivalent to Boolean OR) gate and the output will be 1(6V) for all combinations except the initial. The output from OR gate will be fed to the Fuzzy XOR gates whose other inputs are taken from the outputs of the previous state(Q(t)). As a result the XOR output will produce complement of Q s where ever the OR outputs are 1(according to the basic Boolean logic that is consistent with the Fuzzy). Thus except the initial state, for all states J and K inputs are fed by the complements of Q(t) and this will bring the stability and consistency of output for all successive states. The circuit drawn is applied for RESET TYPE Flip Flop but this same method can be applied for the SET TYPE as well. This can be somewhat mixed circuit where Clock transition detector and it's connected counter will behave according to Boolean algebra but without any contradiction with the FUZZY rules. A major aberration in this proposed Fuzzy Flip Flop except in case of initial state is : The output will be toggled in the next state for J=0 and K=0 and will remain unchanged for J=1 and K=1 and other Fuzzy combinations will behave exactly as formulated in the

previous part of this paper. The mathematical formulation for the proposed RESET type Fuzzy Flip Flop is:

For t=0

$$Q(t+1)=\max\{\min(j,(1-Q(t))),\min((1-k),Q(t))\}$$

For t>0

$$Q(t+1)= \max\{\min\{J,Q(t)\},\min\{(1-K),(1-Q(t))\}\}$$

And for the proposed RESET type Fuzzy Flip Flop is:

For t=0

$$Q(t+1) =\min\{ \max\{J,Q(t)\},\max\{(1-K),(1-Q(t))\}\}$$

For t>0

$$Q(t+1) =\min\{ \max\{(1-J),(1-Q(t))\},\max\{K,Q(t)\}\}$$

This problem may be resolved by simply connecting a Fuzzy inverter with the line that feeds back to the input gates for the next period. By doing this, the table becomes:

Initial Q	J	K	Q(t)	Q(t+1)	Q(t+2)
0	4	4	2	2	2
2	4	4	2	2	2
4	4	4	4	4	4
6	4	4	4	4	4

This solution yields some interesting results:

- i) When J and K are set to 4V(.67 Fuzzy Value) each then upto 2V(.33 Fuzy value) of the initial state output, the outputs in all successive states will be stable and fixed to 2V(.33 Fuzy value) after the initial state . When the output of initial state is 4V(.67 Fuzzy Value) or more then the outputs in all successive states will be stable and fixed to 4V(.67 Fuzzy value)
- ii) There are two cases of No Change situation :
When the output of the initial state is 2V or 4V then the output of all states(including the initial state) will remain unchanged. Thus with .67 fuzzy value for the inputs we can obtain no change situation whereas in Boolean based Flip Flops we get this only when both J and K are 0. This clearly points to the diverse range of outcomes of Fuzzy Flip Flops.

• Sabyasachi Chakraborty is currently working as Sr.Lecturer(IT/CS) and Head of the Department (IT/CS) in Icfai University-Sikkim-India and currently pursuing Ph.D in Application of Fuzzy Logic in Computer Science and Engineering from West Bengal University of Technology-India Email: scit75@gmail.com,Ph:919635107041. The author has performed and guided over 15 software engineering and research projects.

3 CONCLUSION

4 This paper analyses the research work "Non-Associative Fuzzy Flip-Flop with Dual SET-RESET Feature" and reflects a study on Fuzzification of Boolean based flip-flop operations and their dual SET-RESET features. Basically the study and review is objected to explain the formulation of expressions for output of a J-K flip-flop where SET type and RESET type are found to be equivalent. At the end it has been observed that some logic rules are required to be formulated for Fuzzy J-K flip-flop those are applicable both for "SET type" as well as "RESET type" output equations. It is worth noting that the output will be at equilibrium when both J and K are at equilibrium and the output(of next period) are driven by diverse conditions when J and K deviates from equilibrium. Further research can be carried out in this area to incorporate Fuzzy into Boolean based sequential circuit to broaden the variety of output signals so that the Fuzzy enabled circuit can process any hazy value represented by some variations in voltages in the physical level. In this paper firstly the rule base is derived in this paper which may represent various output at diverse input combinations in a well formatted way. Secondly some measure is suggested to avoid the unstable situations arise in different states of Fuzzy J-K Flip Flop. There is a huge scope of further research to resolve the problem of said unstable situation. With the diverse range of outcomes, the Fuzzy Flip Flops brings much broader dimensions in sequential processing as compared to the Boolean based Flip Flops.

Hungarian-Japanese Symposium SISY 2006 • 4th Serbian-Hungarian Joint Symposium on Intelligent Systems 299on Fuzzy Systems and Applications (Technical University, Budapest, 1991), pp. 60-63

[3] J. C. Fodor, M. Roubens: Fuzzy Preference Modeling and Multicriteria Decision Support, Kluwer Academic Publishers, Dordrecht 1994, p. 272

[4] I. R. Goodman: A Decision Aid for Nodes in Command& Control Systems based on Cognitive Probability Logic, Proceedings of the 1999 Command& Control Research and Technology Symposium, US Naval War College, Newport, RI, 1999, pp. 898-941

[5] K. Hirota, K. Ozawa: Concept of Fuzzy Flip-Flop, Preprints of 2nd IFSA Congress, Tokyo, 1987, pp. 556-559

[6] K. Hirota, K. Ozawa: Fuzzy Flip-Flop as a Basis of Fuzzy Memory Modules, In: M. M. Gupta et al., eds., Fuzzy Computing. Theory, Hardware and Applications, North Holland, Amsterdam, 1988, pp. 173-183

ACKNOWLEDGMENT

I, as the author want to thank the top management of Icfai University Sikkim for providing with continuous encouragement, support and motivation to carry out the Research Work. I will remain indebted to my PhD guides from West Bengal University of Technology for providing me with enormous support and showing me the proper directions.

REFERENCES

[1] B. Bouchon-Meunier, V. Kreinovi

ch, H. T. Nguyen: Non-Associative Operations, Second International Conference on Intelligent Technologies (InTech 2001), Bangkok, Thailand

[2] J. C. Fodor, L. T. Kóczy: Some Remarks on Fuzzy Flip-Flops, In: L. T. Kóczy, K. Hirota, eds., Proc. of the Joint

Security Attacks in Wireless Sensor Network

Prachi Bansal, Beenu Yadav, Sonika Gill, Harsh Verma

Abstract— Wireless Sensor Networks (WSNs) use small nodes with constrained capabilities to sense, collect, and disseminate information in many types of applications. One of the major challenges wireless sensor networks face today is security. Wireless Sensor Networks (WSN) is an emerging technology and have great potential to be employed in critical situations like battlefields and commercial applications such as building, traffic surveillance, habitat monitoring and smart homes and many more scenarios. One of the major challenges wireless sensor networks face today is security. In this paper we present an introduction to wireless sensor networks, its usage in every environment followed by a brief overview of characteristics and requirements for deploying such a network. The different attacks on these networks are discussed. For each of these attacks, counter measures are presented if applicable.

Index Terms— Authentication, Confidentiality, DoS (Denial of Service), Routing, Security Goal, Security Attacks, WSN (Wireless Sensor Network).

I. INTRODUCTION

WSN are composed of a large set (hundreds to a few thousand) of homogeneous nodes with extreme resource constraints [1]. Each sensor node has wireless communication capability plus some level of intelligence for signal processing and data networking. These nodes are usually scattered over the area to be monitored to collect data, process it, and forward it to a central node for further processing. Military sensor networks might detect and gather information about enemy movements of people and equipment, or other phenomena of interest such as the presence of chemical, biological, nuclear, radiological, explosive materials. In almost every environment different kinds of sensors are in use. Sensors are used in buildings automation for controlling lights, access control, refrigeration control or HVAC control. Industrial automation uses different kinds of sensors such as sensors for temperature sensing and control, pressure sensing, level sensing and machinery monitoring. Power and utility automation use sensors for remote reading of residential meters or for power distribution diagnostics. Environmentalist uses them for environmental monitoring to measure air and water quality as well as seismic activity, health specialist uses them for tele-health monitoring and diagnostics where they significantly reduce overall medical costs by enabling home-based proactive monitoring and medical care, like personalized patient-based monitoring techniques for measuring the heart rate or respiration. sensors can also be used for maintaining the integrity and safety of buildings, industrial facilities, roadways, water supplies and other public infrastructure. In short – different kinds of sensors are used in our day to day environment to detect, monitor, collect data obtained in different environments.

Nowadays several different wire-based or actuators network products can be found in building automation, industrial automation, security systems or automotive systems. Wired sensors though of great use share some; namely are expensive to install, inflexible once installed, limited in size, in complexity, in functionality and are highly obtrusive in existing infrastructure. On the other hand wireless sensor networks are not restricted by these limitations.

Wireless sensor networks offer advantages in terms of scalability for multi-hop networks with ten to thousands of devices, robustness because of self reconfigurations and distributed intelligence, profitability through low installation costs and flexibility in terms of wireless data collections [1], [3]. The application of their operation includes building controls (fire alarms systems), thermo technology climate control systems, in military for tracking and monitoring borders and so on. This shows that the content of transmitted data covers a spectrum of applications from low security like thermo technology to the high security requirement for military purposes.

Depending on the kind of application it might be necessary to transmit information to other parties. Data collected by outside temperature sensors might be forwarded to a computing system which uses the information for weather forecast applications. The security system, which for example detects an alarm, might inform the owner directly by email or an SMS about the alarm in his house, besides contacting security facilities.

It must be ensured that a central unit is able to collect the data and information about the wireless sensor network. Therefore, a connection from the wireless sensor network to this central unit is required. Hence the challenge that lies with wireless sensor networks is the security of data transmission, reduce power consumption and cost reduction.

Figure 1, shows the layout of a sensor networks where sensor nodes are used to collect the data which is passed through a transit network through multi –hops to reach the base station where the processing of data takes place and then it is forwarded to data service center for storage and analysis of the data collected takes place [4].

-
- Prachi Bansal, Lecturer, Department of CSE, Teerthanker Mahaveer University, Moradabad, India, E-mail: enggprachi@gmail.com
 - Beenu Yadav, Assistant Professor, Department of CS, College of Professional Education, Meerut, India, E-mail: beenu_yadav@rediffmail.com
 - Sonika Gill, Lecturer, Department of CSE, K N G D Modi Engg. College, Modinagar, India, E-mail: sonika.gill108@gmail.com
 - Harsh Verma, Assistant Professor, Trident Group Of Institutions, Gazia- bad, India, E-mail: er.harsh86@gmail.com

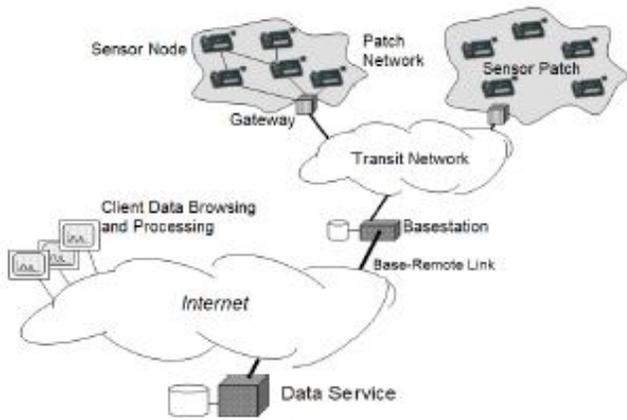


Fig. 1: Layout of Wireless Sensor Networks [2]

A. Layout of Node

Figure II (a): shows a typical layout of a sensor node which comprises of a power unit which is the battery, a processing unit which consist of the memory and embedded processor (Tiny OS which is mostly used as a operating system) , a sensing unit the sensor and a communication unit that is the transceiver.

With the flow of information dictating the criticality of these applications, it is pertinent to secure these networks from malicious or destructive entities and threats.

This motivates the need of security for Sensor Networks. Therefore we look into various threats that could hamper the integrity of this network

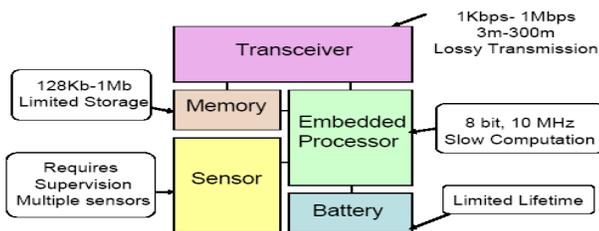


Figure II (a) : Layout of a node [3]

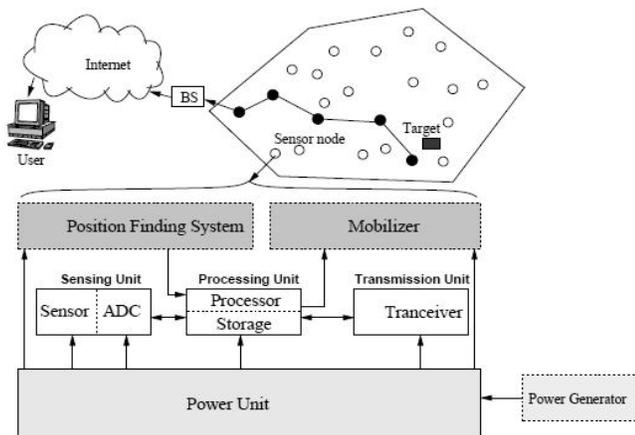


Figure II(b) : Components of a Sensor node

B. Characteristics and requirements

In this section, the different physical characteristics are discussed. Knowing these will help in understanding the difficulties in implement a truly secured solution in wireless sensor networks.

- *Small in size and low power consumption* [2]: Wireless sensors or nodes are small in size so that they can be placed in any environment such as to monitor fire alarms, road traffic, forests, oceans, etc. Being small adds to the ease of use and also brings up the issue of power consumption. These devices are made to utilize low power for computation, processing and data transfer to enable energy efficiency. As it is unfeasible to recharge thousands of nodes every month or in weeks.
- *Concurrency-intensive operation* [2]: The prime goal of wireless sensors is to allow data flow within the network with minimum amount of processing at each node. The communication is established by communicating through multi-hop in order for the data to reach the base station.
- *Diversity in design and usage* [2]: Networked sensors are devised to be application specific as opposed to general purpose, because of their size and function. Therefore minimum requirements are met at both hardware and software level.
- *Low cost* [2]: Each scenario whether it be for military purposes, or track weather conditions, hundreds and thousands of nodes are used in each case to form a network; hence the cost should be low for their deployment.
- *Security* [2], [7]: Secure networks need to be devised for maintaining the integrity of data. A lot of research is being carried out in this field to enable proper deployments of secure wireless sensor networks.

II. SECURITY GOALS FOR SENSOR NETWORKS

As the sensor networks can also operate in an adhoc manner the security goals cover both those of the traditional networks and goals suited to the unique constraints of adhoc sensor networks. The security goals are classified as primary and secondary . The primary goals are known as standard security goals such as Confidentiality, Integrity, Authentication and Availability (CIAA). The secondary goals are Data Freshness, Self-Organization, Time Synchronization and Secure Localization [7].

The primary goals are:

A. Data Confidentiality

Confidentiality is the ability to conceal messages from a passive attacker so that any message communicated via the sensor network remains confidential. This is the most important issue in network security. A sensor node should not reveal its data to the neighbors.

B. Data Authentication

Authentication ensures the reliability of the message by identifying its origin. Attacks in sensor networks do not just involve the alteration of packets; adversaries can also inject additional

false packets. Data authentication verifies the identity of the senders and receivers. Data authentication is achieved through symmetric or asymmetric mechanisms where sending and receiving nodes share secret keys. Due to the wireless nature of the media and the unattended nature of sensor networks, it is extremely challenging to ensure authentication.

C. Data Integrity

Data integrity in sensor networks is needed to ensure the reliability of the data and refers to the ability to confirm that a message has not been tampered with, altered or changed. Even if the network has confidentiality measures, there is still a possibility that the data integrity has been compromised by alterations. The integrity of the network will be in trouble when:

- A malicious node present in the network injects false data.
- Unstable conditions due to wireless channel cause damage or loss of data.

D. Data Availability

Availability determines whether a node has the ability to use the resources and whether the network is available for the messages to communicate. However, failure of the base station or cluster leader's availability will eventually threaten the entire sensor network. Thus availability is of primary importance for maintaining an operational network.

The Secondary goals are:

E. Data Freshness

Even if confidentiality and data integrity are assured, there is a need to ensure the freshness of each message. Informally, data freshness suggests that the data is recent, and it ensures that no old messages have been replayed. To solve this problem a nonce, or another time related counter, can be added into the packet to ensure data freshness.

F. Self-Organization

A wireless sensor network is a typically an ad hoc network, which requires every sensor node be independent and flexible enough to be self-organizing and self-healing according to different situations. There is no fixed infrastructure available for the purpose of network management in a sensor network. This inherent feature brings a great challenge to wireless sensor network security. If self-organization is lacking in a sensor network, the damage resulting from an attack or even the risky environment may be devastating.

G. Time Synchronization

Most sensor network applications rely on some form of time synchronization. Furthermore, sensors may wish to compute the end-to-end delay of a packet as it travels between two pairwise sensors. A more collaborative sensor network may require group synchronization for tracking applications.

H. Secure Localization

Often, the utility of a sensor network will rely on its ability to

accurately and automatically locate each sensor in the network. A sensor network designed to locate faults will need accurate location information in order to pinpoint the location of a fault. Unfortunately, an attacker can easily manipulate non-secured location information by reporting false signal strengths, replaying signals.

This Section has discussed about the security goals that are widely available for wireless sensor networks and the next section explains about the attacks that commonly occur on wireless sensor networks.

III. SECURITY ATTACKS

In this section, the different security attacks of the wireless sensor network are discussed. We begin by investigating the denial of service attacks that are usually implemented on the physical layers. Then, various different tactics that are target at the routing mechanism of WSNs are discussed. Some of these attacks also applies to general ad-hoc wireless networks.

A. Denial of Service Attacks

Due to the importance of functionality of sensors in WSNs (Wireless Sensor Networks), it is imperative that the resources of the network be available in order for these sensors to perform their functions. Denying service to these sensors by any means will severely debilitate the functioning of not only the sensors, but also the effected areas of the network. This issue is not only limited to denial of resources, but also extends to flooding of resources in the network to render it useless during the period of the attack.

Sensor networks could be vulnerable to these types of attacks at different layers of the protocol stack.

1) Physical Layer DoS Attacks

The physical layer of sensor networks comprises of a wireless medium which is universally accessible without requirement of any physical connectivity. This renders these networks more susceptible to attacks like jamming of medium and tampering of physical nodes causing denial of service.

- *Jamming*. An adversary can jam radio frequencies used by sensor nodes. This could affect part or the entire sensor network depending on how wide the attack is. Though jamming can be easily sensed by the network, counter measures are often resource heavy resulting in effecting the functioning of the network [5]. Effective counter measures include use of spread-spectrum communication, switching to lower duty cycles to conserve power, isolating jammed region to circumvent communications, etc.
- *Tampering*. Tampering of nodes involves physically interrogating nodes to obtain valuable data or information like cryptographic keys in order to be used to gain access to higher level communications [5]. Though may not be feasible to limit access to hundreds of nodes in a network, they can be built tamper-resistant to some extent. Nodes should react to illegal

interrogation in ways such as auto-erasing critical information so as to not compromise the same.

2) Link Layer DoS Attacks

Denial of service attacks at the link layer could include creating collisions, or exhausting nodes using very minimal resources.

- *Collision*. Collisions can be created by subverted nodes using minimal amount of energy [5]. These attacks could cause the networks involvement in expensive back offs resulting in failure to perform functions. Error correcting codes could provide a counter measure against these attacks; however, they need the full cooperation of all nodes in the network.
- *Exhaustion*. A subverted node could continually invoke requests to communicate with other nodes are provoke responses, thereby exhausting the limited power resources of the node [5]. Rate limiting could effectively work around these types of attacks.

3) Network Layer DoS Attacks

The network layer plays a critical role in sensor networks. In the absence of routing infrastructure in these networks, all nodes are expected to route vital information at some point or the other. Attacks on protocols in this layer can severely debilitate the functioning of the networks. Some of the vulnerabilities are discussed below [5]:

- *Neglect and Greed*. Subverted or malicious nodes could participate in exchange of some data/information between neighbors but choose to drop important information, not routing it to its intended destination. On the other hand, a malicious node could broadcast itself as the shortest route to the destination, thereby attracting all traffic creating congestion around it. Counter measures include using multiple routing paths and/or redundant messages.
- *Homing*. This involves targeting nodes that have special responsibilities in a locality. As these nodes provide critical services, they are likely candidates for attack. Once subverted, location and presence of critical resources is divulged.
- *Misdirection*. This is a more active attack in which malicious nodes can misdirect traffic along wrong paths by advertising wrong routes. Only authorized nodes exchanging routing information can be an effective counter measure against this type of attack.
- *Black Holes*. Malicious nodes in networks using distance vector routing protocols can advertising zero-cost routes to other nodes. This results in all nodes directing traffic to this adversary node, thereby exhausting resources of the neighboring nodes creating a black hole. Again, authorized exchange of routing information can be a deterrent to this type of attack.

4) Transport Layer DoS Attacks

Attacks on protocols in this layer aim at disrupting reliable communication between two sensors. Types of DoS

attack could be [5]:

- *Flooding*. Adversaries can send several connection establishment requests, thereby causing sensors to allocate memory to maintain the connections. A counter measure against this attack is solving client puzzles.
- *De-synchronization*. Malicious nodes can disrupt reliable communication between two nodes by forging messages to the sender and/or receiver, thereby rendering their communication useless. Authentication of each message could counter this attack.

B. Spoofed, Altered or Replayed Routing Information

Routing information exchanged between sensors can be falsified or altered by malicious nodes to attract traffic towards itself. The same can also be done to ward traffic off important routes. Also, routing information can be replayed to loop information in circles amongst the same nodes exhausting their vital energy resources.

Counter measures against these types of attacks is authentication amongst nodes that route traffic during exchange of routing information. This would prevent malicious nodes from establishing themselves within the network and inadvertently getting neighboring nodes to route information to them [6].

C. Selective Forwarding

Routing in sensor networks relies on cooperation of each intermediate node to dedicatedly route all information directed towards it to the next hop. Adversaries can exploit this situation by subverting a node on a path of major data flow and selectively forward only some messages to the next hop.

A counter measure to this is to use redundant routes (multi-path routing) to pass on information. In case one route is compromised due to an adversary, redundant messages can reach the destination and pass on messages.

D. Sinkhole Attacks

Compromised nodes are made attractive to other nodes in the region by advertising incorrect routing information or high quality routes. This prompts most nodes in the area to route their traffic through this subverted node. This malicious node has now created a 'sink' in the region and is now handling a lot of traffic crucial to the network.

Now that the subverted node has created a 'sphere of influence' attracting traffic to be routed through it, it can perform different types of actions like selective suppression of packets or data modification on information sent from any node in the region.

Notice that sinkhole attacks are particularly of interest in these types of networks as all data is routed to one final destination [6].

Counter Measure

One way to overcome this issue is to implement a hierarchical system of routing information where each region

has a leader node and the leader nodes forward information to the base station. If the different nodes switch performing the role of leaders in their regions, it would prevent subversion for a prolonged duration of time.

Another way to work around this attack is to use geographic routing protocols. These protocols use local routing information and dynamically establish routes to the base station. Hence, attraction towards sink holes is minimal.

E. Sybil Attack

In a Sybil Attack, an adversary node assumes multiple identities, thus presenting itself to the network as multiple nodes [6]. This could cause ineffectiveness in a network, especially the ones that implement fault tolerant schemes and ones that uses geographic routing protocols.

Counter Measure

A protocol can be created to counter Sybil attacks. In such protocol, each node is assigned one or more "verified" neighbors. The base station also sets the number of neighbors a node is allowed to have. A node is allowed to route its data through anyone of its neighbors (verified or not). However, the base station keeps track of how many neighbors each node has. If a node has more than the specified upper limit (an indication of a possible Sybil attack), an error message is sent to that node, and it is then only allowed to communicate through its verified neighbors. Neighbor verification can be implemented through digital certificates or any public key crypto system.

F. Wormhole Attack

A wormhole attack is one in which an adversary node tunnels messages from one part of the network to another, usually through a low latency link [6]. This attack is usually performed using two powerful adversary nodes, located at different side of the network, in order to attract traffic. This attack is usually used in conjunction with selective forwarding or eavesdropping.

Counter Measure

Wormhole attacks are hard to detect, simply because in most cases, the communication medium or protocol between the two adversary nodes are unknown. The only way to detect or counter wormhole attacks is to somehow control and verify the hop counts for each message received by the base station. However, this scheme severely limits the self-organizing criteria of an ad-hoc network. Also, it is possible for the adversary nodes to mimic hop counts by altering the routed messages.

Wormhole attack takes advantage of the fact a route is calculated based on hop counts. Hence, one can design a protocol that doesn't use hop counts, thus deeming the attack meaningless. For example, in a geographic routing protocol, a route is created based on the coordinates of the sending node and intermediate nodes. Unless an adversary node can mimic its location, it is hard for it to attract traffic.

G. HELLO Flood Attack

In many protocols, when a new sensor node is introduced into the network, it broadcasts a HELLO message. Any nodes that can hear the message will reply. Through this mechanism, the new node identifies its neighbors, and also let the neighbors know of its existences. If the new node provides a better route to the base station, its neighbors will change their existing routes such that data is now routed through the new node.

In a HELLO flood attack, a powerful adversary node whose transmission range is farther than a typical node is placed in the network [6]. It broadcasts the HELLO message, advertising a very high quality base station link. Any nodes receiving this message will reply, thinking that a better route can be created through this adversary node. Additionally, due to the high power transmission of the message, it also reaches the nodes that are outside the normal range of a typical node. The target nodes attempt to reply to the adversary node. However, the replies are sent in vain – the target nodes are not able to send messages strong enough to reach their destination.

Counter Measure

The solution for this attack is quite simple. The HELLO protocol can be extended to a three-way handshake [6]. In this protocol, the new node broadcasts a HELLO message. Any receiving nodes that can hear the message send a nonce to new node. The new node must resend the nonce back to each receiving node. The receiving nodes verify the reply with the original nonce to confirm the link. Hence, this protocol guarantees the bi-directionality of a link before any meaningful messages can be sent through it, thus countering the HELLO flood attack.

H. Acknowledgement Spoofing

In most sensor network (and any network for that matter), an acknowledgment is sent by the receiver back to the sender to confirm the safe arrival of the data. However, due to the broadcast medium used by most sensor networks, an adversary node can easily intercept message sent between two parties.

An acknowledgement spoofing attack is one in which an adversary node spoofs an acknowledgement to the sender, even through the message might not be properly received [6]. The goal of this attack is to convince the sender that a weak link is strong or a dead or disabled one is still active.

IV. CONCLUSION

Due to their communication layers and routing topologies, wireless sensor networks are vulnerable to many different kinds of attacks, many of these are applicable also to general ad-hoc network. Unlike ad-hoc networks, these attacks are complicated by the physical limitations of wireless sensors. Finding suitable solutions for each of these attacks is indeed a very challenging task. However, it is very important that these security problems be solved, knowing that this type of net-

work will soon be applied to military, defense and biological surveillance applications.

REFERENCES

- [1] Vierira, Marcos and Silva, Diogenes. "Survey on Wireless Sensor Network Devices". Technical report *IEEE-0-7803-7937/03, Federal University of Minas Gerais*, 2003
- [2] Hill, Jason and Culler, David. "System Architecture Directions for Networked Sensors". Technical report *ACM 0-89791-88-6/97/05*, University of California, Berkeley, 2004
- [3] Archana Bharathidasan, Vijay Anand Sai Ponduru. "Sensor Networks: An Overview". University of California, Davis.
- [4] Culpepper, B.J., Dung, L., Moh, M., "Design and Analysis of Hybrid Indirect Transmissions (HIT) for Data Gathering in Wireless Microsensor Networks". *ACM Mobile Computing and Communications Review (Mc2R)*, Jan/Feb 2004.
- [5] Wood, D. Anthony and Stankovic, A. John. "Denial of Service in Sensor Networks". Technical Report *IEEE-0018-9162/02, University of Virginia*, 2002.
- [6] Chris Karlof, David Wagner. "Secure Routing in Wireless Sensor Networks: Attacks and Countermeasures". University of California at Berkeley.
- [7] P.K. Goel, V.K. Sharma, "Wireless Sensor Network: Security Model", *International Journal of Science Technology and Management, IJSTM Vol. 2, Issue 2, ISSN: 2229-6646 (online)*, pp 100-107.

Insight of Lectins- A review

Himansha Singh¹, Sai Partha Sarathi²

¹School of Translational Medicine, University of Manchester, Oxford Road, Manchester, UK M13 9PT. Contact at Himansha.singh@gmail.com

²School of Life Sciences, University of Salford, The Crescent, UK, M5 4WT

Abstract— Lectins are the carbohydrate-binding proteins and are known for their toxic relevance. History provides various evidences for their toxicity profile and initially, it was thought that lectins are associated solely with toxic content. However, recent research shows a remarkable development in lectin science and their use in studying glycoprotein biosynthesis, investigating carbohydrates on cells/cell-organelles, mapping neuronal pathways, anti-cancer therapies and in mitogenic stimulation of lymphocytes is well documented. Even though lectins have intrigued current researchers for the investigation of their therapeutic values, not enough studies have been conducted so far and knowledge about lectins is limited to certain plants or animal sources and further research is important to identify lectins and their importance in as many sources as possible. This review is a comprehensive account of lectins starting from their initial discovery, their purification, biological role and also presents an outline of their toxic effects and therapeutic contribution. In the search for their functions, the end may not be in sight but, at last, it is conceivably around the corner. Lectins could be the next generation medicines if efficient research is contributed in their understanding.

Index Terms— Lectins, History, Proteins, Toxin, Purification, Future Therapeutics, Biological role of lectins, Carbohydrate specificity

1.0 LECTINS- AN APPRAISAL

According to the widely accepted definition by Goldstein, "A Lectin is a carbohydrate-binding protein or glycoprotein of non-immune origin which agglutinates or precipitates glyco-conjugates or both" (1). It means that the lectins are assumed to be multivalent and the specificity is largely dependent on monosaccharide termini. Some plant lectins may agglutinate various blood groups of erythrocytes and are therefore, called phytohemagglutinins (1). In brief, Lectins are carbohydrate-binding proteins that are found in most plants, particularly seeds and tubers such as cereal crops, potatoes, and beans (legumes). Current research has shown the extraction of lectins from animal sources as well (2). Traditionally, they have been used as histology and blood transfusion reagents. Lectins may be toxic, inflammatory, resistant to cooking or digestive enzymes, and are found in much of our foods (3). Even though, lectins are associated with toxicity, their importance in terms of cancer therapeutics, immunology, antibacterial property etc. cannot be neglected. Scientific research has shown that few of the peculiar characteristics of plants and animals (anti-bacterial properties, cytotoxicity etc.) are because of the presence of lectins and the understanding of these characteristics at molecular level could only be gained by thorough investigation of their associated lectins.

2.0 HISTORY

Seeds of certain plants and animals being toxic to mankind have been known for a long time (4). During the latter part of 19th Century, when the science of bacteriology formed the genesis of its understanding and had a marked influence on the scientific thinking and its approach, it was believed that the toxicity of such seeds was due to bacterial toxin. However this theory was critically disproved and was widely discarded by Wander and Waddell; they investigated that the toxicity of jequirity bean, *Arbus precatorius*, resided in a 'fraction', which could be precipitated by alcohol from an aqueous extract of

the bean (4). Several years later, Dixson obtained a highly toxic concentrate from extract of the Castor bean, *Ricinus Communis*. It was established that the toxic element was present in the extract of the seeds. Stillmark was the first to experimentally prove that a protein fraction of the castor bean, which he called '*Ricin*' was capable of agglutinating red blood cells and he termed it as *Phytohemagglutinins*. The work of Stillmark intrigued the attention of Ehrlich, who decided to work with '*Ricin*' rather than the bacterial toxins, which were so popular among the bacteriologists of that time. The use of these substances led Ehrlich to the discovery of the most fundamental principles of immunology (5,6). Few years later significant work was done by Landsteiner, who stated that the relative Hemagglutinating activity of various seed extracts were quite different when tested with erythrocytes from different animals, he also compared this specificity with that of antibodies of animal blood serum (7). The specificity towards specific erythrocytes was further investigated by Boyd and Shapleigh, it was them, who coined the term 'Lectin', derived from the 'latin' word 'legere', to pick, choose, or select (8). Lectins are also referred to, in the literature, as agglutinins or phytoagglutinins (9-11). The contributions incorporated by 'Stillmark' marked the beginning of centennial on lectin identification, purification, characterization and biological properties and functions. Until 1970s, not much was known about lectins, as only few of them were isolated (mostly from plants and few invertebrates) (12).

Over the years, numerous lectins have been isolated from plants as well as from microorganisms and animals, and during the past two decades the structures of hundreds of them have been established. Concurrently, it has also been shown that lectins function as recognition molecules in cell-molecule and cell-cell interactions in a variety of biological systems (13). Table 1 presents a brief account of 100-plus years of lectin re-

search and show how these proteins have become the focus of intense interest for biologists and in particular for the glyco-biologists among them.

Table 1 History of Lectins

Year	Scientist	Contribution
1884	Warden & Waddel	Toxicity in <i>Abrus precatorius</i> seed extracts
1886	Dixson	Toxicity in <i>Ricinus communis</i> seed extracts
1888	Stillmark	Hemagglutinating activity in <i>Ricinus communis</i> seed extracts Toxicity in <i>Croton triglium</i> seed extracts
1890	Power & Cambier	Toxicity in <i>Robinia pseudoacacia</i> seed extracts
1890	Erlich	Use of abrin an ricin in immunological research
1891	Hellin	Hemagglutinating activity in <i>Abrus precatorius</i> seed extracts
1893	Siegel	Hemagglutinating activity in <i>Jatropha curcas</i> seed extracts
1897	Elfstrand	Hemagglutinating activity in <i>Croton triglium</i> seed extracts
1899	Camus	Hemagglutinating activity in <i>Helix pomatia</i>
1902	Landsteiner	Reversibility of the hemagglutination by heat
1902	Kauss	Inhibition of the hemagglutinating activity by non-immune serum
1903	Noguchi	Hemagglutinating activity in horseshoe crab
1907	Landsteiner & Raubitschek	Hemagglutinating activity in non-toxic plants
1908	Wienhaus	Agglutination of leucocytes and kidney and liver cells by <i>Phaseolus vulgaris</i>
1908	Landsteiner & Raubitschek	Species specificity of plant hemagglutinins
1909	Mendel	Hemagglutinating activity in <i>Robinia pseudoacacia</i> seed extracts
1909	Landsteiner	Inhibition of hemagglutinating activity by heat treated serum
1909	Landsteiner & Raubitschek	Inhibition of the heagglutinating activity by mucin
1912	Schneider	Hemagglutinins and germination
1917	Sumner	Isolation and crystallization of Concanavalina A (Con a)
19126-7	Marcusson-Begun/Siever	Aplicability of lectins for blood typing
1935	Sugishita	Specificity of eel serum agglutinins
1936	Sumner & Howell	Sugar specificity of Concanavalin A
1947-9	Boyd & Reguera/Renkonen	Blood group specificity of plant heagglutinins

1949	Liener	Toxicity of <i>Phaseolus vulgaris</i> hemagglutinins
1949	Jaffe	Thermoinactivation of <i>Phaseolus vulgaris</i> hemagglutinins
1952	Watkins & Morgan	Inhibition of lectins by simple sugars
1954	Boyd & Sharpleigh	Introduction of the term lectin
1960	Nowell	Mitogenic stimulation of lymphocytes by <i>Phaseolus vulgaris</i> lectin
1963	Aub	Agglutination of malignant cells by lectins
1964	Muclenaere	Parallel inactivation of hemagglutinating and antinutritional activity by heat
1965	Agrawal & Goldstein	Affinity chromatography for lectin purification
1966	Boyd	Lectins in algae
1970	Apsberg <i>et al.</i>	Use of Con A for affinity purification of glycoproteins
1974	Ashwell & Morel	Role of animal lectins in endocytosis of glycoproteins
1976	Gallo	Interleukin 2 dissolved in medium of lectin stimulated lymphocytes
1977	Ofek <i>et al.</i>	Role of bacterial lectins in infection
1980	Pusztai	Interaction of <i>Phaseolus vulgaris</i> lectin with intestinal wall
1981	Reisner <i>et al.</i>	Use of lectins in bone marrow transplantation
1984	Yajko <i>et al.</i>	Combined use of lectin and enzyme in clinical identification of micro-organism
1987	Harban-Mendoza <i>et al.</i>	Control of root-knot nematodes by lectins
1988	De Oliveira <i>et al.</i>	Lectin and pancreas hypertrophy
1989	Diaz <i>et al.</i>	Root lectin as a specificity determinant in the <i>Rhizobium</i> -legume symbiosis
1990	Yamauchi & Minamikawa	Con A expression in <i>Escherichia coli</i> cells

Until the early 1970s, although the presence of lectins was reported in numerous organisms, especially in plants, only few of them had been purified, and almost all the purification was performed by conventional techniques such as salt-induced crystallization, ethanol precipitation, ion exchange chromatography and gel filtration (13). Some of the early-purified lectins include plant lectins from soya beans, green peas, *Dolichos biflorus* seeds, wheat germ, and mushroom (*Agaricus campestris*) (12-13), and animal lectins of eel (13-14), and snail (2). These conventional methods relied on the physicochemical properties of the proteins for separation.

The introduction of Affinity chromatography for the purification of lectins was a turning point in the field of lectinology and immensely increased the pace of purification of lectins from various sources (15-16, 30). Affinity chromatography depends on the specific interaction between the lectin and a carbohydrate structure attached to an inert matrix (30). This breakthrough discovery led the easy availability of numerous lectins for a small period of time. However, even then, plants remained the primary source for the supply of lectins. Although the occurrence of lectins in animals was noted quite early, almost in all invertebrates and in lower vertebrates, only the aforementioned three animals (eel, snail, and horseshoe crab) were actually isolated and characterised. The first lectin from the animal source, shown to be specific for a sugar (L-fucose), was from the eel (17).

Lectins are found in almost all edible plants and exposure of men and animals to lectins is inevitable. The majority of plant lectins are present in seed cotyledons, where they are found in cytoplasm or they may also be present in protein bodies (18-19). Some lectins such as 'ricin' and 'abrin' are highly toxic; the general public became aware of toxicity of ricin after the murder of 'Georgi Markov' a Bulgarian writer. Even during World War II, ricin bomb was developed and tested in British military, but was never used as bio-weapons (13).

3.0 Carbohydrate Specificity

Based on the amino acid sequences of available lectins, it is deduced that the carbohydrate binding property of most lectins resides in a polypeptide sequence, which is termed as 'Carbohydrate-recognition domain' (20). Generally, lectins are classified in five groups on the basis of their affinity for (i) Glucose/Mannose; (ii) Galactose and N-acetyl-D-galactosamine; (iii) N-acetylglucosamine; (iv) L-fucose, and (v) Sialic acid (11). Source and specificities of lectins are summarised in Table 2. The binding with simple or complex carbohydrate conjugates is reversible and non-covalent. The specificity of lectins towards carbohydrates can be defined on the basis of 'Hapten inhibition test', in which various sugars or saccharides are tested for their capacity to inhibit the property of hemagglutination of erythrocytes. The binding property of many lectins can be affected by more than one carbohydrate moiety; this is because each lectin molecule possesses two or more carbohydrate-binding sites that are essential for their

ability to agglutinate cells or to react with complex carbohydrates (22). Hapten inhibition test is commonly used for identification of lectin specificity in present days.

Table 2 Sources and specificity of few of the common lectins (Gal, Galactose; Fuc = fucose; Glc = glucose; GalNAc, N-acetylgalactosamine; GlcNAc, N-acetylglucosamine; Man, mannose) (21)

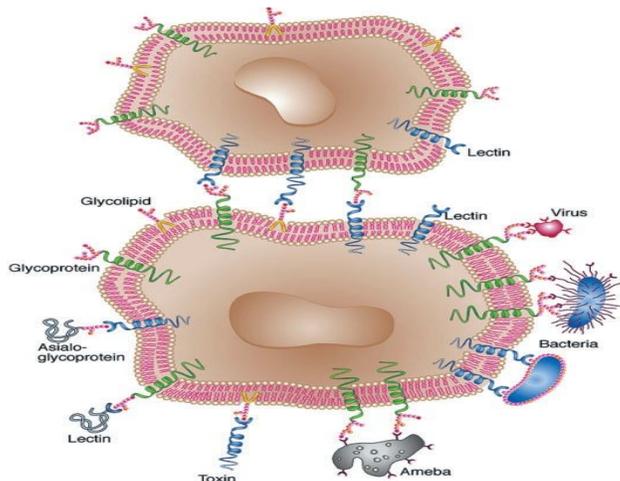
Systemic Name of the plant source	Common Name	Preferred Abbreviation	Specificity
<i>Aaptos papilleta</i>	Sponge	AAP	GlcN Ac
<i>Abrus precatorius</i>	Jequirty bean	APA	Gal, GalNAc
<i>Aegapodium podagraria</i>	Ground elder	APP	Complex
<i>Agaricus bisporus</i>	Common mushroom	ABA	Complex
<i>Albizzia julibrissin</i>	Mimosa tree seed	ALJ	Nonspecific
<i>Allomyrina dichotoma</i>	Japanese beetle	AlloA	Gal, GlcAc
<i>Aloe arborescens</i>	Aloe plant	AAR	Nonspecific
<i>Amphicarpaea bracteata</i>	Hog peanut	AMB	GalNAc
<i>Anguilla Anguilla</i>	Eel	AAA	Fuc
<i>Aplysia depilans</i>	Molluse from Mediterranean sea	AGL	Galacturonic acidGal
<i>Arachis hypogaea</i>	Peanut	PNA	Gal, GalNH ₂
<i>Artocarpus heterophyllus</i>	Jacalin	JCA	Complex
<i>Bauhinia purpurea</i>	Camel's foot tree	BPA	GalNAc, Gal
<i>Bryonia dioica</i>	White bryony	BDA	Complex
<i>Canavalia ensiformis</i>	Jack Bean	Con A	Man, Glc
<i>Caragana Arborescens</i>	Siberian pea tree	CGA	GalN Ac
<i>Carcinocorpius rotundacauda</i>	Horseshoe crab	CCN	Sialic Acid

4.0 Biological Role

The biological role of lectins is based on conjecture rather than knowledge. The question of the possible physiological role of lectins has intrigued investigators from the beginning and the main highlight has been given to plant derived lectins, which for a long time were virtually the only ones known (23). It has been reported that feeding bruchid beetles with a diet containing 'the black bean' lectin resulted in the death of the bruchid larvae (24). On this basis, scientists concluded that the major role of lectins in legumes could be related to protection from attack by insect seed predators. In addition, it was found

that lectins may be involved in protection of plants against pathogenic microorganisms as well based on the observation that showed that WGA, PNA, and SBA inhibited the sporulation and growth of fungi such as *Trichoderma viride*, *Penicillium notatum*, and *Aspergillus niger* (25). Researchers found that the major carbohydrate specificity groups of about 11 lectins were all found to cause growth disruption during germination of spores of *Neurospora crassa*, *Aspergillus amstelodami*, and *Botryodiplodia theobromae* (26). Furthermore, it is speculated that lectins are involved in sugar transport or carbohydrate storage, because of their role in adhesion and agglutination. Lectins have also been considered to be important in both symbiotic as well as in pathogenic interaction between some microorganisms and host. They also play important role in microbial adhesion to various surfaces; they can bind to mucosal membrane and resist denaturation by acid as well as by proteolytic enzymes (27-28). Lectins are also shown responsible for the specific association between nitrogen-fixing rhizobia and leguminous plants, which provides the plant with the needed nitrogen, was advanced nearly three decades ago (29). It was based on the theory that lectin from a particular legume was bound to the surface polysaccharide or lipopolysaccharide of the corresponding rhizobial species, but not to bacteria that are symbionts of other legumes. The suggestion has therefore been made that rhizobial attachment to plant roots occurs by interaction between the bacterial surface carbohydrates and lectins present in the roots of the leguminous plants. This is known as the lectin recognition hypothesis and still the subject of controversy, because of the lack of unequivocal evidence. Lectins serve as means of attachment of different kinds of cells as well as of viruses to other cells via the surface carbohydrates of the cells to be attached (28). In some cases, cell surface lectins bind to particular glycoproteins, whereas in other cases the carbohydrates of cell surface glycoproteins or glycolipids serve as sites of attachment for biologically active molecules that have specificity towards carbohydrate, for example, microorganisms, various plant toxins, galactin etc., as shown in Fig 1.

Figure 1 Specificity of cell surface carbohydrates towards various biomolecules (28)



The significance of lectins and its function in recognition or cell surface interaction occurred after 1950's, and it was demonstrated that influenza hemagglutinin is responsible for the attachment of the virus to the host cell prior to infection (13). Few years later, research showed the ability of lectins to provide innate immunity in animals. For instance, urinary tract infection in mice by mannose specific *Escherichia coli* can be prevented by methyl α -D-mannoside; several other lectins have also been proved to provide innate immunity. A prominent example is the mannose specific receptor present on the surface of macrophages, this receptor later binds to the infectious organism that expose mannose-containing glycans on their surface, resulting in ingestion and killing of the foreign organism (13). A recently discovered lectin of this type is dectin-1, specific for β 1, 3 and/or β 1, 6-glucans, present on fungi. Several lines of research have shown the importance of animal lectins as well, in marking essential biological properties, such as Calnexin, calreticulin, ERGIC-53, Collectins, Dectin-1, Galectins, Macrophage mannose receptor, Selectins to name a few (31). They have been involved in defence mechanism, lymphocytes homing and interactions of immunological cells. They also find a prominent role in cell biology i.e. in cell-cell interactions, cell growth, apoptosis, cell division and cell cycle (See Table 3).

Table 3 Functions of Lectins (31)

Lectin	Role in
Microorganisms	
· Amoeba	Infection
· Bacteria	Infection
· Influenza virus	Infection
Plants	
· Various	Defense
· Legumes	Symbiosis with nitrogen-fixing bacteria
Animals	
· Calnexin, calreticulin, ERGIC-53	Control of glycoprotein biosynthesis
· Collectins	Innate immunity
· Dectin-1	Innate immunity
· Galectins	Regulation of cell growth and apoptosis; regulation of the cell cycle; modulation of cell-cell and cell-substratum interactions
· Macrophage mannose receptor	Innate immunity; clearance of sulphated glycoprotein hormones
· Man-6-P receptors	Targeting of lysosomal enzymes
· L-selectin	Lymphocyte homing
· E- and P-selectins	Leukocyte trafficking to sites of inflammation
· Siglecs	Cell-cell interactions in the immune and neural system

5.0 Lectins- Toxins or Therapeutics

Toxic relevance of lectins is known since ages. Some lectins are resistant to the gut enzymes and do not break easily, they might bind to the wall of the gut and cause damage to the gut lining. This could be related to diseases such as colitis, Crohn's disease, Coeliac-Sprue, and IBS (3). A study by P.G Brady showed that Lectins were active in the faeces of rats and human beings, when the subjects were fed on food sources rich with these proteins. Brady and his colleagues isolated wheat germ agglutinin, a plant lectin by affinity chromatography and also isolated, purified and identified wheat germ agglutinin from faecal samples (32). Research shows that lectin derived amino acids are non-available for animals and the partially digested lectins can bind to the epithelial cell lining of the intestines (33-34). Jaffe was the first to attribute poor performance of rats after the ingestion of *Phaseolus vulgaris* (kidney beans lectins) (35). Years of research has shown that after the interaction of lectin with the intestine, it is edocyted and cause many disturbances in the systemic levels (35). Few similar experiments were performed by De oliveria and his colleagues, they concluded that when lectin derived from pure *Phaseolus vulgaris* were fed to rats, enlargement of the intestine, liver, and pancreas was observed. Apparently, this enlargement in pancreas may be responsible for the observed decrease in the insulin levels of the rats (36). It was also observed that the kidney bean lectin fed rats had thymus atrophy, it is speculated that this atrophy was developed may be because of the unusual proliferation of bacteria in the gut, as the immunological system of the rats may have been depressed due to toxic effects of lectins (36). The ingestion of kidney bean lectin also disturbs the intermediary metabolism, leading to loss of weight, inadequate development and eventually death of the experimented rats (36). Other diseases associated with lectins are insulin dependent diabetes, rheumatoid arthritis, IgA nephropathy and peptic ulcers. Lectins sensitivity could occur due to the failure of a certain type of barrier protection in the body, namely SIgA barrier protection. It could be argued that lectins are specific to certain carbohydrates and when they attach to their specific carbohydrate substrate, they may damage the cell membrane and may damage the cell (3). Although, lectins possess several toxic elements, their benefits are also documented in the literature making them a subject of therapeutic interest for several researchers

Lectins are used in various biological fields. Their contribution in Cell identification and separation (37), detection, isolation, and structural studies of glycoproteins (38), investigation of carbohydrates on cells and subcellular organelles (31, 43), mapping of neuronal pathways (39) mitogenic stimulation of lymphocytes (40), Purging of bone marrow for transplantation (41) and studies of glycoprotein biosynthesis is remarkable (42).

Certain lectins are also used as carriers for the delivery of chemotherapeutic agents; they are also used in investigating cell surface receptors in various bacteria, protozoa and fungi. Lectins can also be used in determining bacterial cell wall components and bacteriophage receptors (28, 43).

The property of lectins to bind non-covalently to simple sugars and polysaccharides is quite unique, so lectin research has attracted a wide interest in microbial taxonomy (28, 43). Lectins have a role in the clinical laboratory identification and taxonomic classification of many microorganisms. Lectins are generally monoclonal proteins and they possess a spectrum of specificities and molecular weights, due to this, they are classified as substantial tools for diagnostic microbiology applications (28, 43). One of the advantages of applying lectins in clinical microbiology is that cellular or surface receptor sites can be partially analysed by hapten inhibition studies (11, 44). Unlike the production of antisera, which requires pre-treatment of microorganisms for antigen preparation followed by injection of the microorganism and glyconjugate into animals, such as rabbits leading to the absorption of antisera to eliminate nonspecific antibody reactions (45), lectins are simple to use, when they are conjugated to a histochemical label such as fluorescein, peroxidase, or colloidal gold, lectins may be used as histochemical probes to identify and localize specific carbohydrate residues in microorganisms by light or electron microscopy, as well as by blotting methods (45).

The cell binding property of lectins elicit a wide range of biological phenomena for instance, lectins have been used to fractionate animal cells, including B and T lymphocytes and also illustrates changes in cell surface architecture following virus infection or parasite infection (11). They are very important and versatile tools and are applied as probes for fluorescence and electron microscopy as well as in gel diffusion assays. Immobilised lectins are used for affinity chromatography during the separation of glycoproteins as they have advantage over other purification techniques because elution can be pursued with a relatively inexpensive monosaccharide and the glycoprotein, which is to be purified, is not subjected to denaturation (28, 43).

Over the past few years, Lectins have been found to have anticancer properties (46). Several researches have shown the use of Lectins to inhibit tumor growth, especially by causing cytotoxicity, apoptosis and, down-regulation of telomerase activity and inhibition of angiogenesis (46). In addition, lectins have also been found to sequester the pool of available polyamines in the body; thereby help in thwarting cancer cell growth. Some lectins are potent toxins, but due to this ability they could be used as potential therapeutic agents, for instance, lectins such as 'Recin' and 'Abrin' have been coupled to specific monoclonal antibodies and are used in cancer therapy. Ricin could be used for developing specifically cytotoxic chemotherapeutic agents by linking them to cell type-specific monoclonal antibodies (47). Cancer is a deadly disease, where the aberrant behaviour of a single cell type is difficult to treat by chemotherapy (47). It is important in cancer therapy that the treatment targets only the affected cells, leaving the normal cells undisturbed, which is quite difficult, especially in chemotherapy. A promising approach is to incorporate a hybrid reagent that has selectivity towards target cells with potential cytotoxicity. Immunotoxins (ITs) conjugates in which cell reactive monoclonal antibodies are chemically joined to potent toxins are most popular manifestation of such hybrids (48). Target cell specificity is conferred by a monoclonal activity

that has been raised against the tumour cell-specific surface antigen (47). Ricin is the toxin of choice because it is well characterized and has easy purification steps and it is shown that human rarely shows prior immunity to it (49) and therefore, it is one of the most potent toxins known. Other plant lectins, relatively closer in structure and function to ricin are *Abrus pectorius* seeds, *Adenia digitata* roots, *Viscum album* leaves, and *Adenia volkensii* roots (50). Ricin is composed of two distinct N-glycosylated polypeptides joined by disulphide bonds. One of the polypeptide chains (A chain) irreversibly inactivates 60S ribosomal subunits. Ribosomes exposed to A chain cannot bind to EF-2 and hence stops protein synthesis (47). It is stated by Olsnes and Pihl in 1982 that mammalian ribosomes are particularly sensitive to ricin than plant ribosomes. Cell surfaces bind to second polypeptide chain (B chain) of ricin by interaction with galactosyl residues of membrane glycoproteins or glycolipids. Surface-bound ricin then enters the cell by classical receptor mediated endocytosis route, which includes coated pits and vesicles and at some point of time, chain A also translocates across the membrane by Golgi cisternae (51), where it attacks ribosome substrates. Youle and Neville showed that B-chain plays a vital role in transferring toxic chain-A into the cytoplasm (52). Two main ricin based ITs are available; Chain A and B both attached to the cell-specific antibody and only Chain A attached to the cell-specific antibody. Both the forms have their pros and cons and have been widely used in In Vitro clinical applications to deplete tumour growth (47). In conclusion, these researches establish the potential role of lectins as therapeutic and diagnostic tools to combat diseases. Therefore, isolation of these lectins could be beneficial. Further line of research should involve deducing the molecular weight of the lectins. After knowing the molecular weight, amino acid sequencing of the lectin can be executed to generate the RNA coding of the protein that could eventually lead to the identification of the specific cDNA sequence. Such cDNA templates could be saved in cDNA libraries. Using these cDNA libraries, primers of the lectins can be formed to understand their properties and biological role at molecular level. Lectins of therapeutics interests can be preserved and can be used in clinical research and healthcare therapies.

6.0 Conclusion

Although lectins have been marketed as potential toxins, there is now extensive literature reporting the activity of lectins in a number of different tissues and processes, a diversity of reports that illustrates the widespread importance of lectins in cell biology and as a potential therapeutic agent, especially for cancer, molecular/cell biology and immunology are well documented. However, knowledge about lectins is limited to certain plants or animal sources and further research is important to identify lectins and their importance in as many sources as possible. There should be many lines of experimental investigation opening up for exploitation. In the search for their functions, the end may not be in sight but, at last, it is conceivably around the corner. Lectins could be the next generation medicines if efficient research is contributed in their

understanding.

7.0 References

1. Goldstein, I. J., Hughes, R. C., Monsigny, M., Osawa, T. and Sharon, N. (1980) What should be called a lectin? *Nature*, 285, p 66.
2. Hammarström, S. and Kabat, E.A. (1969). Purification and characterization of a blood-group A reactive hemagglutinin from the snail *Helix pomatia* and a study of its combining site. *Biochemistry*, 8, p2696-2705.
3. Natural therapy pages. 2012. Unknown Lectins Available. [ONLINE] Available at: <http://www.naturaltherapypages.co.uk/article/Lectins>. [Accessed 01 January 12].
4. Warden, C. J. H., Waddell, L. A., (1884) *Calcutta Bengal Secretarial Press*, 76.
5. Ehrlich, P., (1897) *Fortschr. Med.* 15, p41.
6. Ehrlich, P., (1891) *Deut. Med. Wochenschr.* 17, p976.
7. Landsteiner, K, 1936. *The Specificity of Serological Reactions*. 3rd ed. Springfield: Charles C Thomas.
8. Boyd, W. C., and Shapleigh, E. (1954) Specific precipitating activity of plant agglutinins (lectins). *Science* 119, p419.
9. Brown, J. C., and Hunt, R. C. (1972) Lectins. *Int. Rev. Cytol.* 52, p277-349.
10. Cushion, M. T., DeStefano, J. A. and Walzer, P. D. (1988). *Pneumocystis carinii*: surface reactive carbohydrates detected by lectin probes. *Exp. Parasitol.* 67, p137-147.
11. Nicolson, G. L. (1974) The interactions of lectins with animal cell surfaces. *Int. Rev. Cytol.* 39, p89-190.
12. Moreira, R., Ainouz, I.L., Oliveira, D., and Cavada, B.S., (1991) *Plant Lectins chemical and biological aspects; Mem.Inst Oswaldo Cruz*, 86, p211-28.
13. Sharon, N., and Lis, H. (2004) History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology*. 14, p53-62.
14. Springer, G.F. and Desai, P.R. (1971) Monosaccharides as specific precipitinogens of eel anti-human blood group H (O) antibody. *Biochemistry*, 10, p3749-3760.
15. Goldstein, I. J., and Hayes, C. E. (1978) The lectins: carbohydrate binding proteins of plants and animals. *Carbohydr. Chem. Biochem.* 35, p601-615.
16. Goldstein, I. J., Hughes, R. C., Monsigny, M., Osawa, T. and Sharon, N. (1980) What should be called a lectin? *Nature*, 285, p 66.
17. Watkins, W.M. and Morgan, W.T.J. (1952) Neutralization of the anti-H agglutinin in eel serum by simple sugars. *Nature*, 169, p825-826.
18. Mialonier, G., Privat, J.P., Monsigny, M., Kohler, G., and Durand, R. (1973) Isolation, properties, physicochemical and localization of *in vivo* lectins of *Phaseolus vulgaris*, (var. rouge). *Physiol Veg* 11, p519-537.
19. Youle, R.J., and Huang, A.H.C. (1976) Protein bodies from the endosperm of castor bean. Subfractionation, protein components. Lectins and changes during germination. *Plant Physiol.* 58, p703-709.
20. Drickamer, K. (1988) Two distinct classes of carbohydrate-recognition domains in animal lectins. *J. Biol. Chem.*, 263,

p9557-9560.

21. Doyle, R. J., Nedjat-Haiem, Miller, F., R. D., and Keller, K. F. (1982) Interaction between plant agglutinins and *Legionella* species. *J. Clin. Microbiol.* 15, p973-975.
22. Wu, A. M., Sugii, S. and Herp, A. (1988) A guide for carbohydrate specificities of lectins. *Adv. Exp. Med. Biol.* 228, p819-847.
23. Etzler, M. E. (1986) Distribution and function of plant lectins, p.371-435. In I. E. Liener, N. Sharon, and I. J. Goldstein (ed.), *The lectins, properties, functions, and applications in biology and medicine.* Academic Press, Inc., Orlando, Fla.
24. Janzen, D.H., Juster, H.B., and Liener I.E. (1976) Insecticidal action of the phytohemagglutinin in black beans on a bruchid beetle. *Science*, 192, p795-796.
25. Barkai-Golan, R, Mirelman, D., and Sharon, N. (1978) Studies on growth inhibition by lectins of *Penicillia* and *Aspergilli*. *Arch. Microbiol.*, 116, p119-121.
26. Brambl, R. and Gade, W. (1985) Plant seed lectins disrupt growth of germinating fungal spores. *Physiol. Plant*, 64, p402-408.
27. Etzler, M.E. (1986) Distribution and function of plant lectins. In Liener, I.E., Sharon, N., and Goldstein, I.J. (Eds.), *the lectins: properties, functions and applications in biology and medicine.* Academic Press, Orlando, FL, pp. 371-435.
28. Lis, H., and Sharon, N. (1986) Application of lectins, p. 293-370. In I. E. Liener, N. Sharon, and I. J. Goldstein (ed.), *the lectins: properties, functions and applications in biology and medicine.* Academic Press, Inc., Orlando, Fla.
29. Bohlool, B.B. and Schmidt, E.L. (1974) Lectins: a possible basis for specificity in the *Rhizobium-legume* root module symbiosis. *Science*, 188, p269-271.
30. Agrawal, B.B.I. and Goldstein, I.J. (1967) Specific binding of concanavalin A to cross-linked dextran gels. *Biochem. J.*, 96, p23C-15C.
31. Lis, H., and Sharon, N. (1989) Lectins as molecules and tools. *Annu. Rev. Biochem.* 55, p35-67.
32. Brady, P. G., Vannier, A. M., and Banwell, J. G. (1978) Identification of the dietary lectin, wheat germ agglutinin, in human intestinal contents. *Gastroenterol*, 75, p236-239.
33. Etzler, M. E. and Branstrator, M. L., (1974) Differential localization of cell surface and secretory components in rat intestinal epithelium by use of lectins. *J. Cell. Biol.*, 62, p329.
34. King, T. P., Pusztai, A. and Clarke, E. M.W., (1980) Immunocytochemical localization of ingested kidney bean (*Phaseolus vulgaris*) lectins in rat gut. *Histochem. J.*, 12, p201-208.
35. Jaffe, W. C., (1960) *Arzmeim-Forsch.*, 10, p1012-1016, quoted by Liener (1986).
36. Oliveira, D., Pusztai, J. T., and Grant, G., (1988) Changes in organs and tissues induce by feeding of purified kidney bean (*Phaseolus vulgaris*) lectins. *I Nutr.Res.* 8, p943.
37. Kobayshi, M., Fitz, L., Ryan, M., Hewick, R.M., Clarck, S.C., Chan, S., Loudon, R., Sherman, F., Perussiaj, I.B., and Trinchierit, G. (1989) Identification and purification of Natural Killer cell stimulatory factor (NKSF), A cytokine with multiple biologic effects on human lymphocytes. *J. Exp. MED.* 170 (1), p827-845.
38. Alroy, J., Orgad, U., Ucci, A.A., and Pereira, M.E. (1984) Identification of glycoprotein storage diseases by lectins: a new diagnostic method. *J Histochem Cytochem.* 32, p1280-1284.
39. Yoshihara, Y., Mizuno, T., Nakahira, N., Kawasaki, M., Watanabe, Y., Kagamiyama, H., Jishage, K.I., Ueda, O., Suzuki, H., Tabuchi, K., Sawamoto, K., Okano, H., Noda, T. and Mori, K.. (1999) A Genetic Approach to Visualization of Multi-synaptic Neural Pathways Using Plant Lectin Transgene. *Neuron*, 22 (1), p33-41.
40. Gary, R., Gunther, J.L., Wang, and Edelman, G.M. (1974) The Kinetics Of cellular commitment during stimulation of the lymphocytes by lectins. *The Journal of Cell Biology.* 62, p366-377.
41. Rhodes, E.G. (1998) Combined lectin/monoclonal antibody purging of bone marrow for use in conjunction with autologous bone marrow transplantation in the treatment of multiple myeloma. *Methods in Molecular Medicine.* 9, p351-61.
42. Wang, Y., Wu, S.L., and Hancock, W.S. (2006) Monitoring of glycoprotein products in cell culture lysates using lectin affinity chromatography and capillary HPLC coupled to electrospray linear ion trap-Fourier transform mass spectrometry (LTQ/FTMS). *Biotechnol. Prog.* 22, p873-80.
43. Sharon, N. and Lis, H. (1989). Lectins as cell recognition molecules. *Science.* 246 (4927), p227-224.
44. Liener, I. E., Sharon, N., and Goldstein, I.J. (ed.). 1986. *The lectins: properties, functions and applications in biology and medicine.* Academic Press, Inc., Orlando, Fla.
45. Slifkin, M., and Cumbie, R., (1987) Identification of group B streptococcal antigen with lectin-bound polystyrene particles. *J. Clin. Microbiol.* 25, p1172-1175.
46. Mody, R., Joshib, S.H.A., Chaney, W. (1995) Use of lectins as diagnostic and therapeutic tools for cancer. *Journal of Pharmacological and Toxicological Methods.* 33 (1), p1-10.
47. Lord, J.M. (1987) The Use of Cytotoxic Plant Lectins in Cancer Therapy. *Plant Physiol.* 85 (1), p1-3.
48. Pastan, I., Willingham, M. C., and FitzGerald, D. J. P. (1986) Immunotoxins. *Cell*, 47, p641-648.
49. Olsnes, S and Pihl, A (1982). *Toxic lectins and related proteins.* Amsterdam: *Molecular Action of Toxins and Viruses*, P. Cohen, and S. van Heyningen. P51-105.
50. Barbieri, L., Stirpe, F. (1986) Ribosome-inactivating proteins up to date. *FEBS Lett* 195, p1-8.
51. Van Deurs B, Tonnessen, T.I., Peterson, O.W., Sandvig, K., Olsen, S. (1986) Routing of internalized ricin and ricin conjugates to the Golgi complex. *J. Cell Biol.* 102 (1), p37-47.
52. Youle, R.J, Neville, D.M., (1982) Kinetics of protein synthesis in activation by ricin-anti-Thy 1.1 monoclonal antibody hybrids. *J Biol Chem*, 257, p1598-1610.

Voltage Stability Improvement using Thyristor Controlled Series Capacitor (TCSC) based on L_{mn} and VCPI Stability Indices

Venu Yarlagadda, Dr.B.V.Sankar Ram, Dr.K.R.M.Rao

Abstract - Reactive power control is the basic requirement for maintaining the voltage levels thereby the stability of the interconnected power system. Voltage variations can be stabilized and controlled by providing required reactive power. These low voltages may also reduce the power Transfer through the transmission lines and may lead to Instability Hence FACTS Controllers are widely used in Interconnected Power Systems to control the voltage levels within the tolerable limits. FACTS Controllers are used to enhance controllability and increase power transfer capability. Among the FACTS devices, the TCSC controller has given the best results in terms of performance and flexibility. An IEEE 9 bus system is programmed in MATLAB and assessed the voltage stability using L and VCPI Stability Indices. Finally the voltage stability has been improved using TCSC

Index Terms— Keywords: TCSC, Voltage Stability, L_{mn} -index, VCP Index, Voltage Stability Improvement. Minimum 7 keywords are mandatory, Keywords should closely reflect the topic and should optimally characterize the paper. Use about four key words or phrases in alphabetical order, separated by commas.

1 INTRODUCTION

Reactive power (vars) is required to maintain the voltage to deliver active power through transmission lines. When there is not enough reactive power, the voltage sags down and it is not possible to push the power demanded by loads through the lines. Many devices contribute to systems reactive power and voltage profile. Example: Capacitors supply reactive power, inductors absorb reactive power.

Voltage stability is required for the perfect functioning of the system; compensating devices are required for the compensation purposes.

1.1 Reactive power

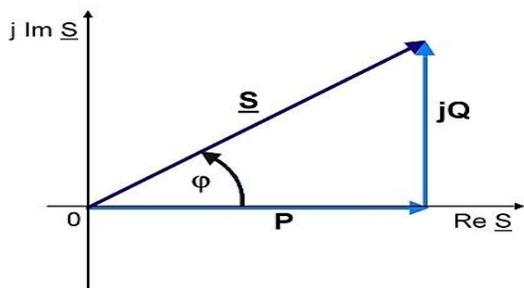


Fig 1. Power Triangle

The figure.1 represents the general phasor diagram showing active power (P) reactive power (Q) where as S represents the apparent power. Reactive power does not transfer energy, so it is represented as the imaginary axis of the vector diagram. Real power moves energy, so it is the real axis.

1.2 Significance of Reactive power

Many devices contribute to reactive power compensation and voltage profile. A transmission line, due to its physical characteristics, supplies reactive power under light loading and consumes it under heavy loading conditions. Power system voltages are controlled through the supply and consumption of reactive power. In general terms, decreasing reactive power margin causes voltage fall, while increasing reactive power margin causes voltage rise. A voltage collapse occurs when the system is trying to serve much more load than the voltage can support.

To maintain efficient transmission and distribution, it is necessary to improve the reactive power balance in a system by controlling the production, absorption, and flow of reactive power at all levels in the system.

1.2 Dependency of Voltage on Reactive Power variations

1.

The Reactive power in a power system has a great deal of impact on the voltage that is prevailing in the system. By compensating the reactive power the voltage profile in the whole power system can be greatly improved which finally leads to the overall improvement of the efficiency and also the power factor of the power system.

For a simple radial Transmission line the relation between receiving end bus voltage is given by, $V_2 = V_1 * Q X / V_1$. In order to keep the receiving end voltage V_2 fixed for a particular sending end voltage V_1 , the drop $(Q X / V_1)$ must remain constant.

- Voltage variations are mostly dependent on reactive power variations
- So, to keep the receiving end voltage constant for constant sending end voltage any deviation of Q must be adjusted locally or by remote control.
- Hence reactive power control is required to maintain the voltage within the acceptable limits.

2 Voltage Instability

Voltage instability is basically caused by an unavailability of reactive power support in an area of the network, where the voltage drops uncontrollable. Lack of reactive power may essentially have two origins: firstly, a gradual increase of power demands without the reactive part being met in some buses or secondly, a sudden change in the network topology redirecting the power flows in such a way that the required reactive power cannot be delivered to some buses. Introducing FACTS devices is the most effective way for utilities to improve the voltage profile and voltage stability margin of the system.

Many power system blackouts all over the world have been reported where the reason for the blackout has been voltage instability.

1) WSCC USA July 2 1996: A short-circuits on a 345 kV line started a chain of events leading to a break-up of the western North American power system. The final reason for the break-up was rapid overload/voltage collapse/angular instability.

2) Florida USA 1985: A brush fire caused the tripping of three 500 kV lines and resulted in voltage collapse in a few seconds.

3 Voltage stability Indices

The purpose of voltage stability indices is to determine the point of voltage instability, the weakest bus in the system and the critical line referred to a bus. These indices are referred either to a bus or a line.

3.1 Line stability Index L_{mn}

This stability criterion is used to find the stability index for each line connected between two bus bars in an interconnected network. This voltage stability criterion is based on a power transmission concept in a single line. Stability criterion is developed considering a single line of a network.



Fig 2. Single Line Diagram of Two Bus Test System

$$L_{mn} = 4XQ_j / [V_i \sin(\theta - \delta)]^2 \quad (1)$$

θ = line impedance angle

δ = angle difference between the sending end and the receiving

end voltage

X = line reactance

Q_j = reactive power flow at the receiving end

V_i = sending end voltage.

The system is said to be stable, in the sense of transmission lines, as long as L_{mn} remains much less than 1; and approaches 1 towards the point of bifurcation. The most critical line connecting the weak buses in the system can be easily identified from the value of L_{mn} closest to 1

3.2 Line stability Index VCPI

The VCPI investigates the stability of each line of the system and they are based on the concept of maximum power transferred through a line.

$$VCPI = P_r / P_{rmax} \quad (2)$$

where the values of P_R e Q_R are obtained from conventional power flows calculations, and $P_R(max)$ and $Q_R(max)$ are the maximum active and reactive power that can be transferred through a line. The VCPI indices varies from 0 (no load condition) to 1 (voltage collapse).

4 Flexible A.C. Transmission System

A Flexible Alternating Current Transmission System (FACTS) is a system comprised of static equipment used for the AC transmission of electrical energy. It is meant to enhance controllability and increase power transfer capability of the network. It is generally a power electronics based device. FACTS devices are defined by the IEEE as "a power electronic based system and other static equipment that provide control of one or more AC transmission system parameters to enhance controllability and increase power transfer capability. FACTS could be connected:

- in series with the power system (series compensation)
- in shunt with the power system (shunt compensation)
- both in series and in shunt with the power system

4.1 Series Compensation

In series compensation, the FACTS are connected in series with the power system. It works as a controllable voltage source

Examples of FACTS for series compensation

Static Synchronous Series Compensator, Thyristor Controlled Series Capacitor, Thyristor Controlled Series Reactor (TCSR), Thyristor Switched Series Capacitor (TSSC), Thyristor Switched Series Reactor (TSSR).

Advantages of FACTS:

- It increases the loading capability of the lines to their thermal capability. Overcoming their limitations and sharing of power among lines can accomplish this.
- Provides greater flexibility in sitting new generation.
- FACTS devices improve the speed of operation of the overall system
- It improves the stability of the system and thus makes the system secure

5 Thyristor Controlled Series Capacitor (TCSC)

The basic conceptual TCSC module comprises a series capacitor, C , in parallel with a thyristor-controlled reactor, L_s , as shown in Fig(a). However, a practical TCSC module also includes protective equipment normally installed with series capacitors.

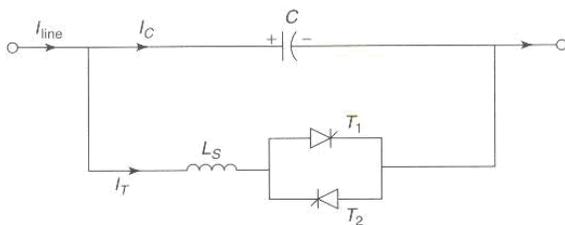


Fig 3. TCSC

Also installed across the capacitor is a circuit breaker, CB, for controlling its insertion in the line. In addition, the CB bypasses the capacitor if severe fault or equipment-malfunction events occur. A current-limiting inductor, L_d is incorporated in the circuit to restrict both the magnitude and the frequency of the capacitor current during the capacitor-bypass operation. A TCSC is a series-controlled capacitive reactance that can provide continuous control of power on the ac line over a wide range.

$$Z_{eq} = -j / [\omega C - (1/\omega L)] \quad (3)$$

If $\omega C - (1/\omega L) > 0$, the reactance of the FC is less than that of the parallel-connected variable reactor and that this combination provides a variable-capacitive reactance are both implied. If $\omega C - (1/\omega L) = 0$, a resonance develops that results in an infinite-capacitive impedance-an obviously unacceptable condition. If, however, $\omega C - (1/\omega L) < 0$, the LC combination provides inductance above the value of the fixed inductor. This situation corresponds to the inductive mode of the TCSC operation.

The behaviour of the TCSC is similar to that of the parallel LC combination. The difference is that the LC-combination analysis is based on the presence of pure sinusoidal voltage and current in the circuit, whereas in the TCSC, because of the voltage and current in the FC and thyristor-controlled reactor (TCR) are not sinusoidal because of thyristor switching.

6 Improvement of System Stability Limit Using TCSC

It is advantageous to install a TCSC in transmission paths, which can adapt its series-compensation level to the instantaneous system requirements and provide a lower loss alternative to fixed-series compensation.

The series compensation provided by the TCSC can be adjusted rapidly ensure specified magnitudes of power flow along designated transmission line. This condition is evident from the TCSC's efficiency, that is, ability to change its power flow as a function of its capacitive-reactance setting:

$$P_{12} = (V_1 V_2 / X_L - X_C) \sin \delta \quad (4)$$

Where P_{12} = the power flow from bus 1 to bus 2 and V_1, V_2 are the voltage magnitudes of buses 1 and 2, respectively

X_L is the line-inductive reactance

X_C is the controlled TCSC reactance combined with fixed-series capacitive Reactance, δ = the difference in the voltage angles of buses 1 and 2.

This change in transmitted power is further accomplished with minimal influence on the voltage of interconnecting buses, as it introduces voltage in quadrature. The freedom to locate a TCSC almost anywhere in line is a significant advantage

6.1 Advantages of TCSC

Use of thyristor control in series capacitors potentially offers the following little-mentioned advantages:

- Rapid, continuous control of the transmission-line series-compensation level.
- Dynamic control of power flow in selected transmission lines within the network to enable optimal power-flow conditions and prevent the loop flow of power.
- Damping of the power swings from local and inter-area oscillations.

7 IEEE-Nine Bus Test System

9 bus system is there as above given figure.4 it contains 9 buses, load, Generators. As per the project undertaken before, TCSC has been placed in the system to improve the voltage stability. The process undertaken for it is the GS method using

the matlab simulation.

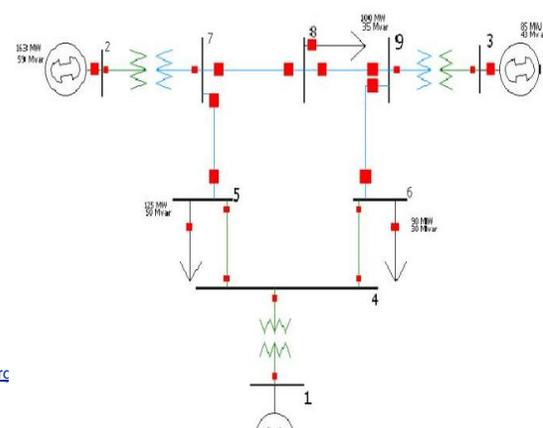


Fig 4. IEEE 9-Bus Test System

The results have been shown with the stability indices and the results are shown as below.

8 Case Study: Results

Table 1 shows the Bus voltages of IEEE 9-Bus Test system with and without TCSC. Similarly Table.2 and Table 3 Shows the values of Lmn and VCPI Indices respectively.

Table 1: Bus Voltages Without and With TCSC

S.No	Bus Voltage	Bus Voltage Without TCSC in P.U	Bus Voltage With TCSC in P.U
1	V1	1.04	1.04
2	V2	1.018	1.019
3	V3	1.012	1.013
4	V4	1.003	1.004
5	V5	0.969	0.971
6	V6	0.982	0.984
7	V7	1.004	1.006
8	V8	0.866	0.937
9	V9	1.006	1.008

Table 2: Line Stability Index Without and With TCSC

S.No	Line Stability Index (L _{mn})	L _{mn} Without TCSC	L _{mn} With TCSC
1	1-4	0.1374	0.1374
2	4-5	0.1098	0.1098
3	4-6	0.0455	0.0455
4	3-9	0.0306	0.0306
5	2-7	0.1062	0.1062
6	7-8	0.5576	0.5576
7	7-5	0.0972	0.0972
8	8-9	0.5530	0.1055
9	9-6	0.0545	0.0545

Table 3: VCP Index Without and With TCSC

S.No	Line Stability Index (L _{mn})	VCPI Without TCSC	VCPI With TCSC
1	1-4	0.148	0.13
2	4-5	0.11	0.1
3	4-6	0.05	0.04
4	3-9	0.04	0.03
5	2-7	0.12	0.11
6	7-8	0.58	0.56
7	7-5	0.14	0.09
8	8-9	0.6	0.11
9	9-6	0.09	0.05

1	1-4	0.148	0.13
2	4-5	0.11	0.1
3	4-6	0.05	0.04
4	3-9	0.04	0.03
5	2-7	0.12	0.11
6	7-8	0.58	0.56
7	7-5	0.14	0.09
8	8-9	0.6	0.11
9	9-6	0.09	0.05

8 Conclusions

This paper shows the effect of TCSC on the Voltage stability Improvement. The voltages, Lmn and VCPI Indices with and without TCSC have been recorded. The results show that the voltage profile is better with TCSC and Lmn and VCPI Indices have been improved tremendously to improve the stability.

References

1. N.G. Hingorani and L. Gyugyi, Understanding FACTS, IEEE Press, New York, USA, 1999
2. R.M. Mathur and R.K. Varma, Thyristor-Based FACTS Controllers for Electrical Transmission Systems, IEEE Press and Wiley Interscience, New York, USA, Feb. 2002.
3. D M Tagare, Reactive Power Management, McGraw-Hill Publication, India,1982.
4. IEEE Power Engg. Society/CIGRE, "FACTS Overview", Publication 95 TP 108, IEEE Press, New York, 1995.
5. P. Kundur, Power System Stability and Control, McGraw-Hill, Inc., New York, 1994.
6. IEEE Power Engineering Society, "FACTS Applications", Publication 96 TP 116-0, IEEE Press, New York, 1996
7. W.A.Lyon, Transient Analysis of Alternating-Current Machinery, Chapter 2, John Wiley, New York, 1954.
8. D.Irwin, Basic Engineering Circuit Analysis. 4th Ed., Mac-milan , New York 1993.
9. Jack R. Smith , Programming the PIC microcontroller. , McGraw-Hill Publication, India, 2005.
10. A.Mohamed, G.B.Jasmon, S.Yusoff "A Static Voltage Collapse Indicator using Line Stability Factors" Journal of Industrial Technology, Vol.7, N1, pp. 73-85, 1989
11. P.Kessel, H.Glavitsch "Estimating the Voltage Stability of a Power System" IEEE, Transactions on Power Delivery, Vol.PWRD-1, N3, July 1986

Acknowledgements

We are immensely thankful to Management and Principal of VNR VJIET for providing R&D lab and other facilities to carryout this work.

AUTHORS BIOGRAPHY



Mr. Venu Yarlagadda, is Obtained his B'Tech and M'Tech Degrees from JNT University, Hyderabad, Andhra Pradesh and also perusing PhD From same University. Currently working as Associate Professor in VNR VJIET, AP, India. Current areas

of Interests are FACTS, Power Systems, Intelligent Systems and Electrical Machines. Published One Research paper in International Journal, Presented Two International and One National Conference papers.

E-Mail : venuyar@gmail.com, venu_y@vnrvjiet.in



Dr.B.V.Sankar Ram, Currently working as a Professor of EEE Dept,JNT.University,Hyderabad, He obtained his PhD from JNTU and M.Tech and B.Tech from O.U, Hyderabad. Published Eight Research papers in International Journals, Presented 18 International and National Conference papers, Currently Guiding Ten PhD's. Current areas of Interests are Power System Reliability, Power Systems, Deregulation and FACTS.



Dr. K.R.M Rao, Currently working as Professor and Head MJCET, Hyderabad. He obtained his PhD From Poland and guided two PhD's and Currently Guiding two PhD's. Current Areas of Interests are Power System Reliability, Electrical Drives, Power Electronics and FACTS

Quality Enhancement of Watermarking System Using Discrete Cosine Transform

Priyanka Sharma

Abstract— The growth of networking boosted the use of information technology to a greater coverage and this leads to the comprehensive overview of current digital watermarking techniques. The purpose of this review paper is to present a new-fragile and non-blind watermarking insertion technique is defined using the DCT domain and a mathematical model is proposed using simulink.

Index Terms—DCT(Discrete Cosine Transform),DWT(Discrete Wavelet Transform), FDCT, Fragile, IDCT, JPEG, Non-Blind, Simulink

1 INTRODUCTION

Image authentication using watermarking is quite a different from steganography and cryptography. In cryptographic authentication, the intention is to protect the communication channel and make sure that the message received is authentic and in Steganography offers an interesting alternative to image integrity and authenticity problem. Because the image data is typically very superfluous, it is possible to slightly modify the image so that we can later check with the right key if the image has been modified and identify the personalized portions. So Steganography and Watermarking are more complementary to each other.

Protection of digital information has attracted a lot of attention during the last few years. The most common method to protect the digital information is watermarking schemes like fragile tamper detection and authentication of visual as well as digital information. In formation of digital libraries the care should be taken to prevent it from the unauthorised, malicious and inadvertent act made to digital libraries. Digital watermarking term is derived from the word "wassermarke" means the marks resembles the possessions of water on paper and defined as a process of adding/embedding an image as a watermark to another image. Digital watermarking is still a extremely young research area with its first academic conference held in 1996. Numerous algorithms have been projected and implemented since then. The multiple watermark method is used to solve the problem of authentication. Recently some multiple watermark algorithms have been proposed like CDMA, DCT, and DWT. The requisite of watermarking technique needs to posses the following distinctiveness such as:

- a) Imperceptibility for hidden information.
- b) Watermark must also be highly healthy.

Unplanned watermarks involve transforms that are commonly applied to images during normal uses such as cropping, resizing, and enhancement etc. A particularly interesting form of unplanned attack is that of **image compression**. In this paper visible watermarking is done with the help of the Discrete Cosine Transform (DCT) domain. DCT is a transform from the spatial domain (the pixel values) to the frequency domain. This new domain consists of cosines with increasing frequency. Generally before this transform is performed, the image is segmented into blocks of for example 8 x 8 pixels. The transform is then performed on the individual blocks. The cosine coefficients are computed for each block. This is done, because when the transform would be performed on the whole image, the memory needed would be very large. This method is therefore a form of block transform coding. The image coding standard JPEG segments the image in blocks of 8 x 8 pixels and then performs a DCT on these blocks.

2 JPEG COMPRESSION AND DECOMPRESSION

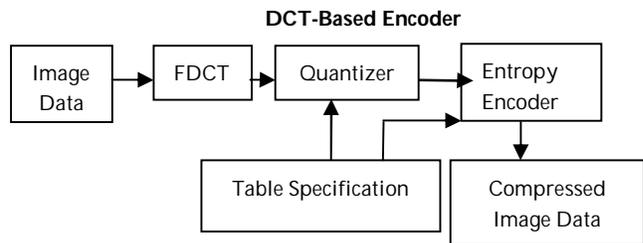


Fig1 DCT based Encoder

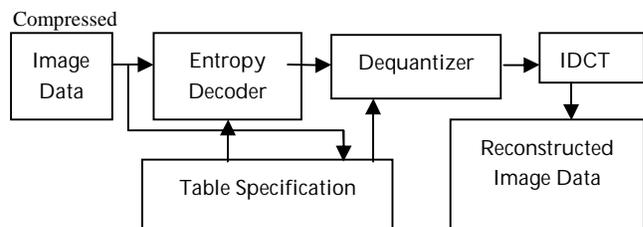


Fig2 DCT based Decoder

JPEG compression steps are discussed below:-

• Priyanka Sharma(M.Tech in Computer Engg) currently working as an Assistant Professor in Computer Science Department of Gurgaon College of Engg, Gurgaon , India, E-mail: priyanka.aqua2008@gmail.com

- 1) The image is first divided into blocks of size 8*8.
- 2) Discrete Cosine Transform (DCT) is the basis for many image and video compression algorithms, specially the baseline JPEG and MPEG standards for compression of still and video images respectively. The equation X The one-dimensional forward discrete Cosine transform (1D FDCT) of N samples is formulated by:-

$$F(u) = \sqrt{\frac{2}{N}} C(u) \sum_{x=0}^{N-1} f(x) \cos\left[\frac{\pi(2x+1)u}{2N}\right] \quad (1)$$

for $u = 0, 1, \dots, N - 1$, where

$$C(u) = \begin{cases} 1/\sqrt{2} & \text{for } u=0; \\ 1 & \text{for otherwise} \end{cases}$$

JPEG compression artifacts blend well into photographs with detailed non-uniform textures, allowing higher compression ratios. The very high compression ratio severely affects the quality of the image, although the overall colors and image form are still recognizable. The function $f(z)$ represents the value of the x^{th} sample of the input signal. $F(u)$ represents a DCT coefficient for $u=0,1,N-1$. The one-dimensional inverse Discrete Cosine transform (1D IDCT) is formulated in a similar fashion as follows:

$$f(x) = \sqrt{\frac{2}{N}} \sum_{u=0}^{N-1} C(u) F(u) \cos\left[\frac{\pi(2x+1)u}{2N}\right] \quad (2)$$

Where $x=0,1,\dots,N-1$.

- 3) The DCT coefficients than quantized using a quantization matrix.
- 4) The quantized coefficients are then arranged in zig zag order.
- 5) Then compressed using the Huffman encoder.

For each block B, the quantized DCT coefficients obtained from the JPEG file are multiplied by the quantization coefficients stored in the quantization matrix. Then the Inverse Discrete cosine transform (IDCT) is computed.

3 WATERMARK IMPLEMENTATION USING MATLAB

Matlab is a tool that was originally designed to simplify the implementation of jpeg image compression. In this paper we make use of matlab Simulink which is the model based approach. We have to simulate our model and analyzes dynamic systems. It enables you to pose a question about a system, model the system, and see what happens. With Simulink, you can easily build models from scratch, or modify existing models to

meet your needs. Simulink supports linear and nonlinear sys

tems, modeled in continuous time, sampled time, or a hybrid of the two. Systems can also be multirate — having different parts that are sampled or updated at different rates. The Matlab Simulink model of watermarking is constructed using ordinary blocks available in video and image processing tool box of Matlab Simulink. The implementation of watermarking system for simulation in Simulink of Matlab is depicted in figure 1. In this model, there is video viewer which views the watermark image.

Model-Based Design is a process that enables faster, more cost-effective development of dynamic systems, including control systems, signal processing, and communications systems. In Model-Based Design, a system model is at the center of the development process, from requirements development, through design, implementation, and testing. The model is an executable specification that is continually refined throughout the development process.

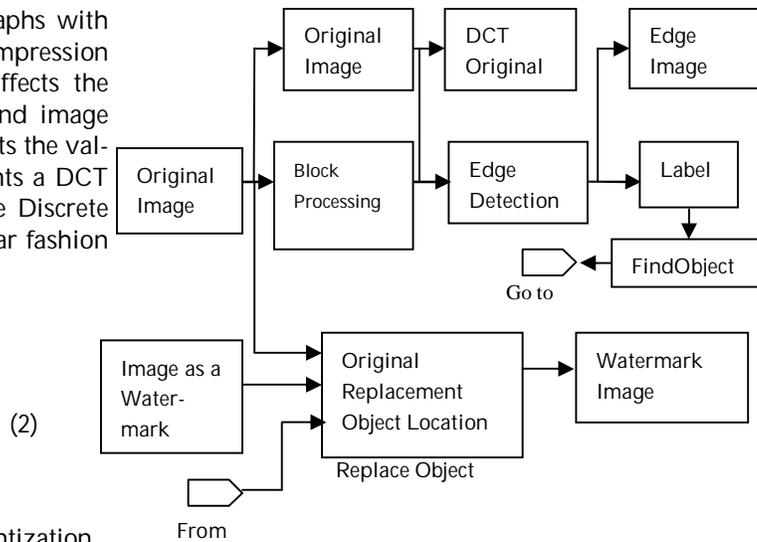


Fig:1 Digital image watermarking DCT(Discrete Cosine Transform)

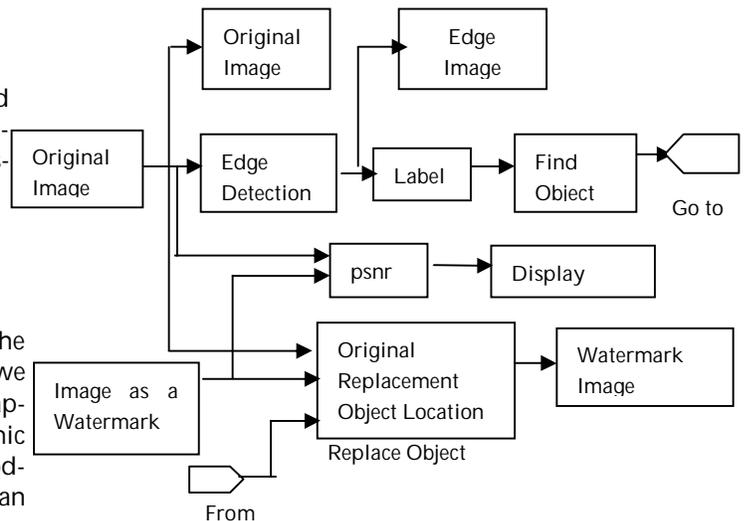


Fig:2 Digital image compression without DCT(Discrete Cosine Transform)

4 RESULT AND CONCLUSION

In this work, analysis and modeling of watermarking using MATLAB Simulink is carried out. We have to analysis of watermarking model with the help of DCT and we also analysis the model without DCT and compare the watermark output of the two by calculating the PSNR value which is shown in Table 1 and we have to conclude that watermark image with DCT is easily stored and take less memory space because compression ratio is high as compare to without DCT watermarking, to find the DCT of image it is necessary to find out the gain matrix of an image and which is find out with help of edge detection, edge means higher amplitude pixel and higher amplitude gives higher gain coefficient. Scanning will be start from higher gain coefficient to lower gain coefficient so that we can easily find out the DCT of an image.

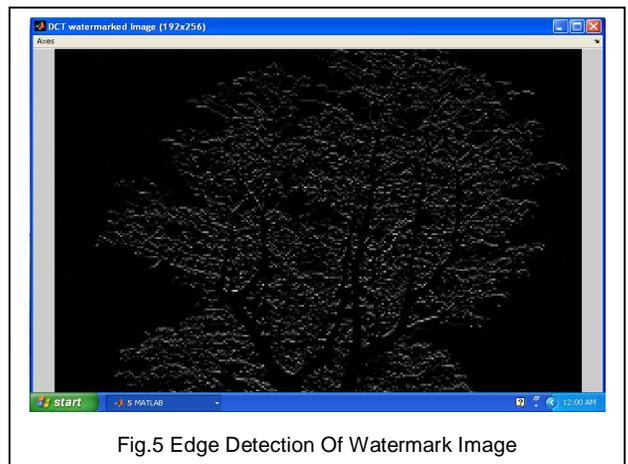
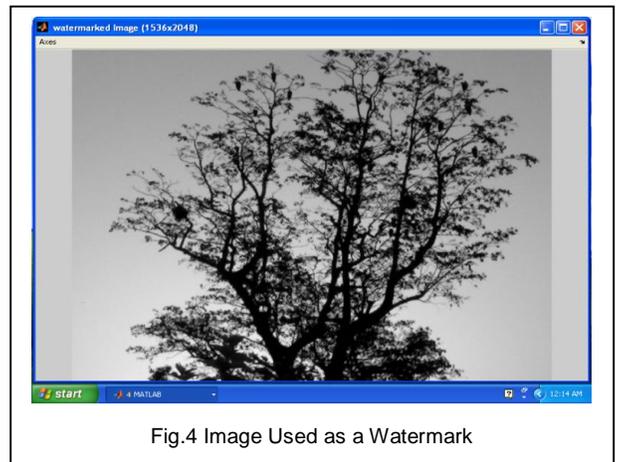
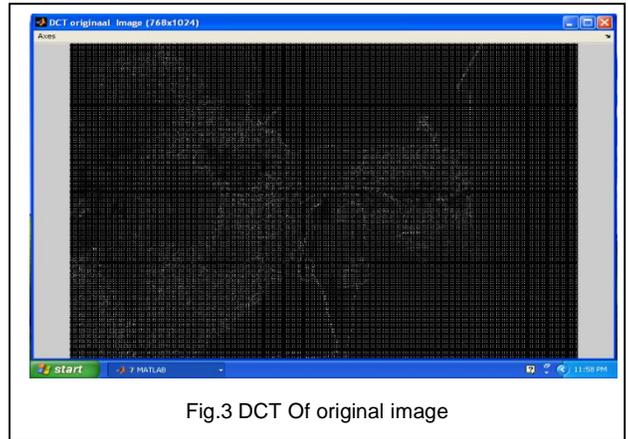
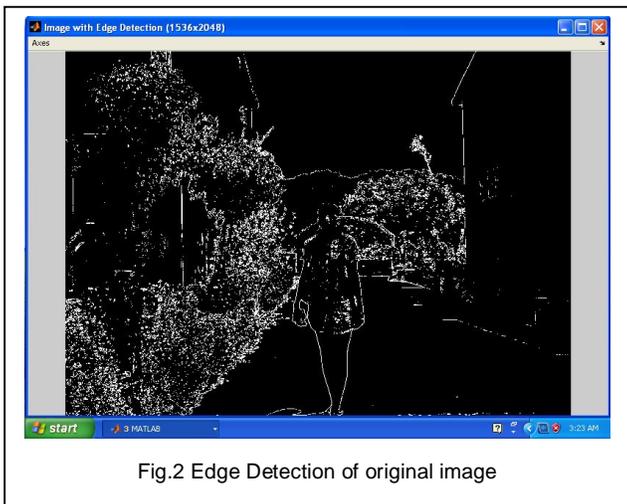
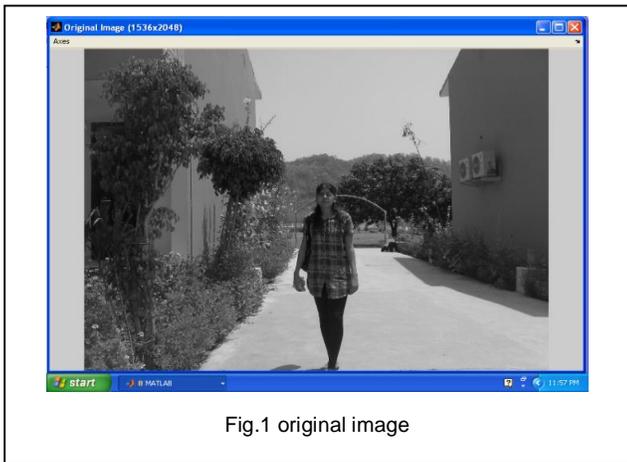




Fig.6 Watermarking of Original Image

TABLE 1

DIFFERENT PSNR VALUE FOR DIFFERENT TECHNIQUES

<u>S.no</u>	<u>Technique</u>	<u>Calculated Value of PSNR</u>
<u>1.</u>	Without Using DCT	7.848
<u>2.</u>	With Using DCT	1.835

5 IMPROVEMENT AND FUTURE WORK

In our experiments, we have noticed that while using DCT some disadvantages is also associated with DCT like the segmentation of the image in blocks. Especially at a high compression ratio, it is possible that the blocks do not connect well to each other so this will overcome by making use of DWT(Discrete Wavelet Transform) because the image is not segmented into blocks deformations. Another improvement of the Wavelet Transform is that the wavelet functions which generally are used, are non-zero on a short interval, in contrary to the cosine functions of the Discrete Cosine Transform.

ACKNOWLEDGMENT

In the end I would like to thank my guide Dr. Rahul Rishi (HOD in CSE Dept. in The Technological Institute Of Textile And Sciences, T.I.T &S, Bhiwani) without whom this work cannot completed and also Mr. Krishan Kumar (Asstt. Proff. In BITS, Bhiwani, ECE Dept).

REFERENCES

- [1] Hsu, C., Wu, J.: Hidden digital watermarks in images IEEE Transactions on Image Processing, 1999,
- [2] Wong, P. Au, O., Yeung, Y.: A novels blind multiple watermarking techniques for images. IEEE Transactions on Circuits and Systems, 2003,
- [3]Raval, M., Priti, P: Discrete wavelet transform based multiple watermarking schemes. In: 2003 Conference on Convergent Technologies for Asia-Pacific Region, 2003.
- [4]Lu, C., Liao, H.: Multipurpose watermarking for im-age authentication and protection. IEEE Transactions on Image Processing, 2001.
- [5]O'Sullivan, J. A., Moulin, P., Ettinger, J. M.: Infor-mation theoretic analysis of steganography. In: Proc.IEEE International Symposium on Information Theory, (1998).
- [6]Cox, I. J., Miller, M. L.: The first 50 years of electronic watermarking. In: EURASIP of Applied Signal Processing, (2002).
- [7]Servetto, S., Podilchuk, C., Ramchandran, K.: Capacity Issues in Digital Image Watermarking. In: International Conference on Image Processing,New York (1998).
- [8] G.K.Wallace . "The JPEG still image compression standard." Communication of the ACM april 1991.
- [9] M.F Brantley and L.P. hurd ,image compression LTD 1993.

Performance Evaluation Of Extended Visual Cryptography Schemes With Embedded Extended Visual Cryptographic Scheme.

Miss. Asha S.N.¹, Dr. Shreedhara .K.S.M.Tech., Ph.D², Smt. Anitha G. BE, ME³,

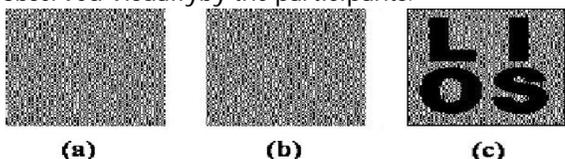
Abstract: Visual cryptography scheme (VCS) is a kind of secret sharing scheme which allows the encoding of a secret image into n shares that distributed to n participants. The beauty of such scheme is that a set of qualified participants is able to recover the secret image without any cryptographic knowledge and computation devices. Extended visual cryptography scheme (EVCS) is a kind of VCS which consists of meaningful shares (compared to the random shares of traditional VCS). Intent of this paper is the study and construction of EVCS which is realized by embedding random shares into meaningful covering shares, and we call it the embedded extended visual cryptography scheme (embedded EVCS). Experimental results compare some of the well-known EVCS's proposed in recent years systematically, and show that the proposed embedded EVCS has competitive visual quality compared with many of the well-known EVCS's in the literature. Besides, it has many specific advantages against these well-known EVCS's respectively.

Keywords: Secret sharing, cryptography, encoding, Embedded, Extended visual cryptography scheme.

I. Introduction

The basic principle of visual cryptography scheme (VCS) was first introduced by Naor and Shamir. VCS is a kind of secret sharing scheme [1, 2] that focuses on sharing secret images. The idea of the visual cryptography model proposed in [3] is to split a secret image into two random shares (printed on transparencies) which separately reveals no information about the secret image other than the size of the secret image. The secret image can be reconstructed by stacking the two shares. The underlying operation of this scheme is logical operation *OR*. In general, a traditional VCS takes a secret image as input, and outputs n shares that satisfy two conditions: (1) any of shares can recover the secret image; (2) Any forbidden subset of shares cannot obtain any information of the secret image other than the size of the secret image. An example of traditional (2, 2)-VCS can be found in the following Figure 1, where, generally speaking, a (k, n) -VCS means any k out of n shares could recover the secret image. In the scheme of Figure 1, shares (a) and (b) are distributed to two participants secretly, and each participant cannot get any information about the secret image, but

after stacking shares (a) and (b), the secret image can be observed visually by the participants.



(a) Figure 1: An example of traditional (2, 2)-VCS with image size 128x128.

Many other applications of VCS, other than its original objective (i.e. sharing secret image), have been found, for

example, authentication and identification [4], watermarking [5] and transmitting passwords [6] etc.

The term of extended visual cryptography scheme (EVCS) was first introduced by Naor et al. in [3], where a simple example of (2, 2)-EVCS was presented. Generally, an EVCS takes a secret image and n original share images as inputs, and outputs n shares that satisfy the following three conditions:

(1) Any qualified subset of shares can recover the secret image; (2) Any forbidden subset of shares cannot obtain any information of the secret image other than the size of the secret image; (3) All the shares are meaningful images. Examples of EVCS can be found in the experimental results of this paper, such as Figures 2-9.

There have been many EVCS's proposed in the literature. Droste [7], Ateniese et al. [8] and Wang et al. [9] proposed three EVCS's, respectively, by manipulating the share matrices. Nakajima et al. [10] proposed a (2, 2)-EVCS for natural images. Tsai et al. [11] proposed a simple EVCS, where its shares were simply generated by replacing the white and black sub-pixels in a traditional VCS share with transparent pixels and pixels from the cover images respectively. Furthermore, Zhou et al. and Wang et al. [12-14] presented an EVCS by using halftoning techniques. Their methods made use of the complementary images to cover the visual information of the share images. Recently, Wang et al. proposed three EVCS's by using error diffusion halftoning technique [15] to obtain nice looking shares. Their first EVCS also made use of complementary shares to cover the visual information of the shares as the way proposed in [12]. Their second EVCS

imported auxiliary black pixels to cover the visual information of the shares. In such way, each qualified participants did not necessarily require a pair of complementary share images. Their third EVCS modified the halftoned share images and imported extra black pixels to cover the visual information of the shares.

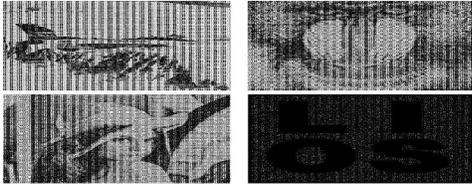


Fig. 2. Shares and the recovered secret image of an embedded (3, 3)-EVCS after reducing the black ratios, the image size is 1024 x 1024.



Fig. 3. Experimental results of (2, 2)-EVCS proposed in [7]-[9]. The size of all the images is 768 x 768.



Fig. 4. Experimental results of (2, 2)-EVCS proposed in [15], [12]. The size of all the images is 768 x 768.



Fig. 5. Experimental results of Method 2 for (2, 2)-EVCS proposed in [13]. The size of all the images is 768 x 768.



Fig. 6. Experimental results of Method 3 for (2, 2)-EVCS proposed in [13]. The size of all the images is 768 x 768.

However, the limitations of these EVCS's mentioned above are obvious. The first limitation is that the pixel expansion is large. The second limitation is the bad visual quality of both the shares and the recovered secret images.



Fig. 7. Proposed (2, 2)-EVCS. The size of all the images is 768 x 768.

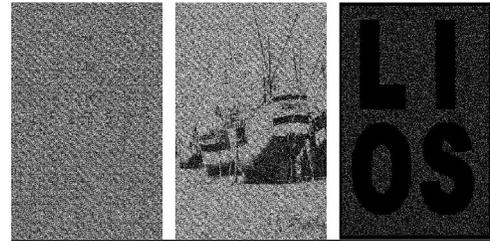


Fig. 8. Experimental results of the second method proposed in [13] for fine share images. The size of all the images is 768 x 768.

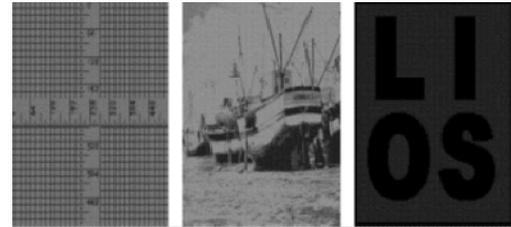


Fig. 9. Proposed (2, 2)-EVCS for fine share images. The size of all the images is 768 x 768.

This paper proposes an embedded EVCS scheme with overall good properties. Comparisons of properties of our proposed scheme with some well-known EVCSs can be found in Section IV, where we will show that our scheme has competitive visual quality compared with many of the well-known EVCSs. Besides, our EVCS has many specific advantages against these well-known EVCSs, respectively. The rest of this paper is organized as follows: In Section II, we introduce the formal definition of embedded EVCS, and give the main idea about our construction. In Section III, we embed the traditional VCS into the covering shares and discuss the bounds of our scheme. Lastly, in Section V, we conclude the paper.

II. A sketch and the main idea of the proposed embedded EVCS

In this section, we will give an overview of our construction. First we introduce the formal definition of embedded EVCS.

Definition 1 (embedded EVCS) Denote Γ^0 and Γ^1 as the basis matrices of a traditional VCS with access structure (Γ, Γ^c) and pixel expansion λ . In order to encode a secret image S , the dealer takes n grey-scale original share images as inputs, and converts them into n covering shares which are divided into blocks of k sub-pixels ($k \geq 1$). By embedding the rows of Γ^0 and Γ^1 (after randomly permuting their columns) into the blocks, the embedded EVCS outputs n shares s_0, \dots, s_{n-1} , and there exist values $\{h : h \in \Gamma\}$, and satisfying:

1. The stacking result of each block of a qualified subset of shares can recover a secret pixel. More precisely, if $\mathcal{I} = \{i_1, \dots, i_t\} \in \Gamma$, denote s_{i_1}, \dots, s_{i_t} as the blocks at the same position of the shares s_{i_1}, \dots, s_{i_t} , then for a white secret pixel, the OR of s_{i_1}, \dots, s_{i_t} is a vector v that satisfies $h \leq v$, and that for a black secret pixel, it satisfies $v \geq h$.

2. Part of the information of the original share images is preserved in the shares. Define $\alpha = (I - S) / I$ be the ratio of the information of the original share images that preserved in the shares, and it satisfies $\alpha > 0$.

In Definition 1, the first condition ensures that the secret image can be visually observed by stacking a qualified subset of shares. The second condition ensures that the shares are all meaningful in the sense that parts of the information of the original share images are preserved. The idea of our embedded EVCS contains two main steps:

- (1) Generate covering shares, denoted as s_0, s_1, \dots, s_{-1} ;
- (2) Generate the embedded shares by embedding the corresponding VCS into the covering shares, denoted as e_0, e_1, \dots, e_{-1} .

III. Embedding the corresponding VCS into the covering shares

Algorithm 1 The embedding process:

Input: The covering shares constructed, the corresponding VCS (s_0, s_1) with pixel expansion M and the secret image I .

Output: The embedded shares e_0, e_1, \dots, e_{-1} .

Step 1: Dividing the covering shares into blocks that contain $(\geq M)$ sub-pixels each.

Step 2: Choose embedding positions in each block in the covering shares.

Step 3: For each black (resp. white) pixel in I , randomly choose a share matrix $\in \{1\}$ (resp. $\in \{0\}$).

Step 4: Embed the sub-pixels of each row of the share matrix into the embedding positions chosen in Step 2.

The diagram of Algorithm 1 can be found in Fig. 11.

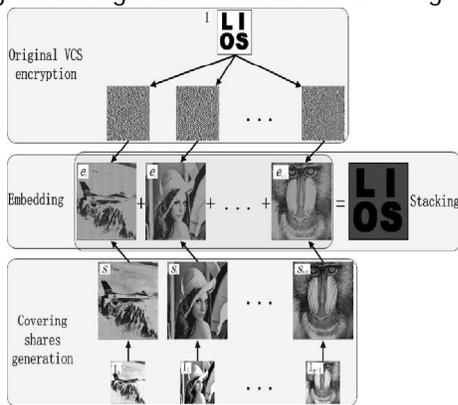


Fig. 11. Diagram of Algorithm 1.

III. Experimental results and comparisons.

In this section, we compare the embedded EVCS with many of the well-known EVCS's in the literature.



Fig. 10. Original share images (airplane, baboon, Lena, ruler, and boat) and the secret image.

First, we give the original images that will be used in the paper (Fig. 10): Lena, airplane, baboon, ruler, boat, and the secret image. The sizes of these images are 256 x 256; they will be scaled to their proper size when necessary.

We provide two well-known objective numerical measurements for the visual quality, the peak signal-to-noise ratio (PSNR) and the universal quality index (UQI) [16]. In this paper, the PSNR is adopted to assess the distortion of each share image with its original halftoned share image (i.e., without the darkening process). In such a way, the PSNR values in Tables IX and X can reflect the effects of a combination of the following possible processes in EVCSs: darkening, embedding, and modification. The PSNR is defined as follows:

$$PSNR = 10 \log_{10} \frac{255^2}{MSE}$$

MSE

Where MSE is the mean squared error. The UQI is adopted to assess the distortion of each share image with its original ray-scale share image (after being scaled to the size of shares). Hence, the UQI value can reflect the effect of the halftoning process besides that of the darkening, embedding and modification processes in EVCSs. The formal definition of UQI can be found in [16]. In this paper, the block size of UQI is set to be 8 for all the experiments.

The original halftoned share images of Zhou *et al.* and Wang *et al.*'s schemes in Figs. 4, 5, 6, and 8 are generated by the blue noise halftoning technique and error diffusion halftoning technique on the original share images in Fig. 10 directly.

TABLE IX
OBJECTIVE NUMERICAL MEASUREMENTS OF FIG. 2

Content	PSNR			UQI			Contrast	PE'	PE''	
	share 1	share 2	share 3	share 1	share 2	share 3				
Fig. 5	No	8.69db	8.62db	8.93db	0.0311	0.0699	0.0215	1/16	16	16

TABLE X

OBJECTIVE NUMERICAL MEASUREMENTS OF FIGS. 3, 4, 5, 6, 7, 8, AND 9

	Content interaction	PSNR		UQI		Contrast	PE'	PE''
		share 1	share 2	share 1	share 2			
Fig. 6	No	3.19db	3.77db	0.0008	0.0032	2/9	9	9
Fig. 7	Yes	9.54db	0.51db	0.0445	-0.0315	1/9*	9	9
Fig. 8	No	3.16db	4.08db	0.0254	0.0304	1/9	9	9
Fig. 9	Yes	4.62db	4.11db	0.0578	0.0332	1/9	9	9
Fig. 10	No	5.67db	6.01db	0.0293	0.0281	1/9	9	9
Fig. 11	No	2.72db	3.61db	0.0438	0.0270	1/9	9	9
Fig. 12	No	3.36db	7.15db	0.0630	0.0289	1/9	9	9

IV. Conclusion

The shares of the proposed scheme are meaningful images, and the stacking of a qualified subset of shares will recover the secret image visually. According to the comparisons with many of the well-known EVCS in the literature [7, 8, 10, 12, 13, 15], the proposed embedded EVCS has many specific advantages against different well-known schemes, such as can deal with grey-scale input images, has smaller pixel expansion, always unconditionally secure, does not require complementary share images, one participant only needs to carry one share and can be applied for general access structure. Furthermore, our construction is flexible in the sense that there exist two trade-offs between the share pixel expansion and the visual quality of the shares and between the secret image pixel expansion and the visual quality of the shares.

Comparisons on the experimental results show that, the visual quality of the share of the proposed embedded EVCS is competitive with that of many of the well-known EVCS's in the literature.

References

[1] A. Shamir. How to share a secret. In *Communications of the ACM*, volume 22 (11), page 612-613, 1979.
 [2] G. R. Blakley. Safeguarding cryptographic keys. In *Proceedings of the National Computer Conference*, volume 48, page 313-317, 1979.
 [3] M. Naor and A. Shamir. Visual cryptography. In *EUROCRYPT '94, Springer-Verlag Berlin*, volume LNCS 950, pages 1-12, 1995.
 [4] M. Naor and B. Pinkas. Visual authentication and identification. In *Crypto '97, Springer-Verlag LNCS*, volume 1294, pages 322-336, 1997.
 [5] T.H. Chen and D.S. Tsai. Owner-customer right protection mechanism using a watermarking scheme and a watermarking protocol.
 [6] P. Tuyls, T. Kevenaar, G.J. Schrijen, T. Staring, and M.V. Dijk. Security displays enabling secure communications.

[7] S. Droste. New results on visual cryptography. In *CRYPTO '96, Springer-Verlag LNCS*, volume 1109, pages 401-415, 1996.
 [8] G. Ateniese, C. Blundo, A. De Santis, and D.R. Stinson. Extended capabilities for visual cryptography. In *ACM Theoretical Computer Science*, volume 250 Issue 1-2, pages 143-161, 2001.
 [9] D.S. Wang, F. Yi, and X.B. Li. On general construction for extended visual cryptography schemes.
 [10] M. Nakajima and Y. Yamaguchi. Extended visual cryptography for natural images. In *WSCG Conference 2002*, pages 303-412, 2002.
 [11] D. S. Tsai, T. Chenc, and G. Horng. On generating meaningful shares in visual secret sharing scheme. In *The Imaging Science Journal*, volume 56, pages 49-55, 2008.
 [12] Z. Zhou, G.R. Arce, and G. Di Crescenzo. Halftone visual cryptography. In *IEEE Transactions on Image Processing*, volume 15, NO.8, pages 2441-2453, 2006.
 [13] Z.M. Wang and G.R. Arce. Halftone visual cryptography through error diffusion. In *IEEE International Conference on Image Processing*, pages 109-112, 2006.
 [14] Z.M. Wang, G.R. Arce, and G. Di Crescenzo. Halftone visual cryptography via direct binary search. In *EUSIPCO '06*, 2006.
 [15] Z.M. Wang, G.R. Arce, and G. Di Crescenzo. Halftone visual cryptography via error diffusion. In *IEEE Transactions on Information Forensics and Security*, volume 4 No.3, pages 383-396, 2009.
 [16] J.O. Limb. Design of dither waveforms for quantized visual signals. In *Bell System Technology Journal*, volume 48,7, pages 2555-2582, 1969.
 [17] G.J. Simmons, W. Jackson, and K. Martin. The geometry of shared secret schemes. In *Bulletin of the ICA*, pages 71-88, 1991.
 [18] Zhou W. and A.C. Bovik. A universal image quality index. In *IEEE Signal Processing Letters*, volume 9(3), pages 81-84, 2002.

Author profile:

Miss. Asha S.N. ¹, Pursuing M.Tech 4th sem, in (CS).

Dr. Shreedhara .K.S.M.Tech., Ph.D²,

Smt. Anitha G. BE, ME³,

Department of Computer Science, University B.D.T. College of Engineering, Davangere-577004, Karnataka Visvesvaraya Technological University, Belgaum, Karnataka - India.
 E mail: ashasn2007@gmail.com. Ph: 09964379359, 09844508359, India.

Design, Modeling and Analysis Of Double Acting Reciprocating Compressor Components

P.K.Jadhao , K.K. Jadhao, S. S. Bhagwat

Abstract : The present day concept of automation has increased the use of compressed air in every field of industrial life. An attempt has been made to design double acting air compressor with an intention to provide the operating pressure required for pneumatic tools within less time taken by single acting compressor to generate the same pressure of 9 kg/cm² at 925 rpm of compressor for piston displacement of 21.4 cfm. Basic component of double acting reciprocating air compressor are designed. The modeling, and analysis of double acting reciprocating air compressor were done by using CATIAV5R10 software. Theoretically, all the components are found to work within safe stress limits..

Key Word : Double Acting Reciprocating Compressor, Stress Analysis, Material Properties, CATIAV5R10, Failure Mode.

1. INTRODUCTION

Compressed air is used in Air refrigeration, cooling of large building for cleaning purposes, blast furnaces, bore wells, spray painting, in super charging IC engines and gas turbines, starting of IC engines, fuel atomizers, compressed air is widely used in braking system of automobiles, railway coaches, wagons etc. and the list is endless where the compressed air is used. In fact today, we find it is extensively used in all fields of application due to Wide availability of fresh air. Compressibility, Easy transportability of compressed air in pressure vessel, containers and long pipes., Fire – proof characteristics of the medium. High degree of controllability of pressure. The detail study of different types of compressor is very much essential. The current study is focused at the study of double acting reciprocating compressors. The advantage of double acting compressor is that it delivers almost double compressed air (almost in half time) which saves time and money of the user. The aim for the current study is to replace the single acting compressor by double acting compressor which generates 9 kg/cm² compressed air on the basis of the tools used in the industries which are generally operated with a maximum pressure of 9 kg/cm². The study is focused on a compressor available in Manoj motor mechanic and servicing center Pusad, which is used for water serving and also for the tyre remolding with the following specifications:

2. OBJECTIVE

The main objectives are to:

- Design Cylinder and Piston for Double acting reciprocating air compressor which generates 9 Kg/cm² compressed air.

- Model and perform stress analysis for Cylinder and Piston .

3. LITERATURE REVIEW

Heinz P Bloch and John J. Hoefner worked on the Development of a Double acting free piston expander for power recovery in transcritical CO₂ cycle.[5]. Sun et al. developed New method of thermodynamic computation for a reciprocating computer simulation by Si – Yieng .[6] W.Norman Shade et.al. suggest optimization and revitalization techniques on compressors used in air drilling, air procession and air separation etc. and emphasis on the fact that virtually any size model can be considered for improvements, A. Al masi worked on reciprocating compressor design and manufacturing with respect to performance, reliability and cost. And suggested methods for optimum reciprocating compressor. A.P.Budagyan and P.I. Platinin devoted on design and optimization on reciprocating compressors [10] and minutely studied the effect of temperature variation on the overall performance of the reciprocating compressors and cooling of compressors. Due consideration is given on optimal basic geometric dimensions of reciprocating compressors .

4. Methodology

4.1 Design of Cylinder and Piston for double acting compressor.

Maximum pressure	= 9 kg/cm ² ,
Diameter of cylinder of piston	93.25 mm
Thickness of cylinder	= 8 mm
Factor of Safety for cylinder	10
Radius of crank	= 75 mm
Thickness of cylinder flange	= 9.6 mm

Size of studs	8
Depth of piston ring	= 4.875 mm
Distance between bottom of piston to bottom of second ring	= 6.42mm
Number of piston rings	2
Thickness of web of piston	= 10 mm
Mass of piston	= 0.9482 kg
RPM of compressor	= 925 rpm
Length of cylinder	= 205 mm
Material of cylinder	= FG 300
Stroke length	= 150 mm
L/D Ratio	1.6
No of studs	4
Thickness of Piston	= 30 mm
Distance between top of piston to top of first ring	= 6.42mm
Clear distance between two piston ring	=7.41 mm
Thickness of piston ring	= 10 mm
Volume of piston	131703.92×10 ⁻⁹ m ³
Outer diameter of piston pin	= 15mm

Material	Aluminum : Alloy 1100-H14
Young Modulus	7.5e+010N_m2
Poisson Ratio	0.346
Density	2800kg_m3
Thermal Expansion	0.0000236
Yield Strength	9.5e+007N_m2

Table 3. Material Properties for Piston

Table 4. Load Computation for Piston

Fx =	2.1144e-014 N
Fy =	1.1577e-014 N
Fz =	-9.6307e+003 N
Mx =	-2.1840e-002 N-m
My =	-4.2288e-003 N-m
Mz =	-1.2020e-019 N-m

5. MODELING AND ANALYSIS OF CYLINDER AND PISTON.

Modeling refers to the process of generating a model as a conceptual representation of some phenomenon. Typically a model will refer only to some aspects of the phenomenon.

Stress computation for cylinder: Table 1 shows the various properties of the material used for cylinder and Table 2 indicate the loading on cylinder Since the maximum pressure inside the cylinder is 0.8829 N/mm²

Table .1 Material properties for cylinder:

Material	Iron : Cast-Iron , Gray 4.5 % , ASTM A-48
Young Modulus	9.239e+010N_m2
Poisson Ratio	0
Density	7800kg_m3
Thermal Expansion	0.0000121
Yield Strength	3.1e+008N_m2

Table .2 Load computation on cylinder:

Fx = -3.137e - 012 N
Fy = 3.125e - 011 N
Fz = -1.207e + 004 N
Mx = 2.956e - 002 N-m
My = 5.285e - 003 N-m
Mz = 7.772e - 016 N-m

Table 5. Forces and reaction for Cylinder

Components	Applied Forces	Reactions	Residual	Relative Magnitude Error
Fx (N)	-3.1370e-012	-5.4353e-012	-8.5723e-012	1.9157e-014
Fy (N)	3.1253e-011	1.0620e-010	1.3745e-010	3.0716e-013
Fz (N)	-1.2067e+004	1.2067e+004	8.1855e-011	1.8292e-013
Mx (Nxm)	2.9564e-002	-2.9564e-002	-1.4809e-011	1.6143e-013
My (Nxm)	5.2850e-003	-5.2850e-003	-1.4206e-011	1.5486e-013
Mz (Nxm)	7.7716e-016	-3.2615e-012	-3.2607e-012	3.5545e-014

Components	Applied Forces	Reactions	Residual	Relative Magnitude Error
Fx (N)	2.1144e-014	-2.6219e-012	-2.6008e-012	6.3020e-014
Fy (N)	1.1577e-014	-2.6312e-013	-2.5155e-013	6.0953e-015
Fz (N)	-9.6307e+003	9.6307e+003	1.8190e-011	4.4077e-013
Mx (Nxm)	-2.1840e-002	2.1840e-002	-1.2267e-013	6.3412e-014
My (Nxm)	-4.2288e-003	4.2288e-003	-3.9167e-013	2.0247e-013
Mz (Nxm)	-1.2020e-019	-1.8957e-014	-1.8957e-014	9.7996e-015

Table 6. Forces and Reactions on Piston

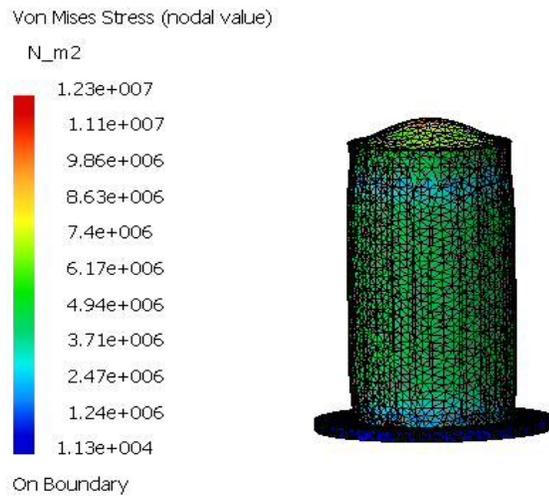


Figure 1. Deformed mesh for cylinder with nodal stresses value

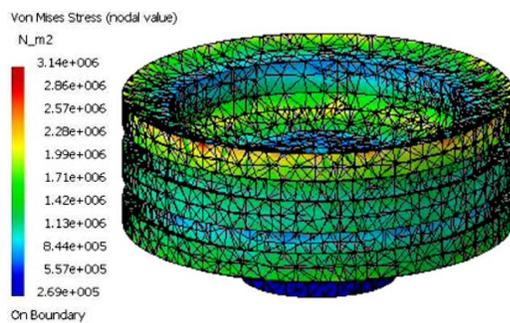


Fig. 2 : Deformed Mesh for Piston with Nodal Stress Values

Entity	Size
Nodes	8547
Elements	35622

For Cylinder

Entity	Size
Nodes	9076
Elements	32627

For Piston

In this an attempt has been made to design a double acting reciprocating air compressor components such as piston and cylinder for maximum pressure of 9Kg/cm² at 925 rpm , for piston displacement of 21.4cfm The fundamental dimensions of each components of double acting compressor were found out analytically and checked for various failures due to induced stresses. The modeling of double acting reciprocating air compressor components were carried out by using CAD software Pro-E Wildfire 4.0 and analysis by using CATIAV5R10 software. Theoretically all the components are found to work within safe stress limits.

BIBLOGRAPHY

- [1] Si – Yieng Sun, Ting – Rong Ren, “New method of thermodynamic computation for a reciprocating computer simulation of working process” International Journal of mechanical science Vol. 37, Issue 4, April 1995 pp 343 – 353.
- [2] Heing P Bloch, John J. Hoefiner, “Design and materials for reciprocating compressor components.” Vol02, pp 32 – 32 April 1992.
- [3] A.P. Budagyan andP.I. Plastinin, “Design and Construction of machines and equipment Optimization of Reciprocating compressor”. *Transited from Khimicheskoe 1 nerflyonoe Mashinostranic, No 3; pp 21 – 22 March, 1981.*
- [4] Mahesh Rathore, *Thermal Engineering*, Tata McGraw Hill Education Private Ltd., New Delhi, 1st Edition 2010, PP. 838 – 863.
- [5] S.R. Majumdar, *Pneumatic System, Principles and Maintenance*, Tata McGraw Hill Publication company Ltd. Edition 1995, PP. 32 – 60.
- [6] Dr. P.C. Sharma and Dr. D.K. Aggarwal, *A Textbook of Machine Design* 11th Edition Publication S.K. Kataria and Son’s, New Delhi, 2006 PP. 842 – 928 and 845
- [7] Joseph Edward Shigley, *Mechanical Engineering Design Internal Edition*, Tata McGraw Hill company, New York 1989. Pp 1210

6. RESULT AND DISCUSSION

Wavelet Filter Banks Modeling of Human Auditory System for Robust Speech Enhancement

Ranganadh Narayanam, Hilmi Dajani and Sos Agaian

Abstract—In this research we have developed a novel perceptual wavelet filter bank architecture & wavelet filter banking coefficients for a versatile speech enhancement method based on the human auditory model. In this paper the implementation of these wavelet filter banking coefficients for a speech enhancement scheme are being described which meets the demand for quality noise reduction algorithms which are capable of operating at a very low signal to noise ratio. This is a generalized time frequency subtraction algorithm which advantageously exploits the wavelet multi-rate signal representation to preserve the critical transient information. This wavelet filter banking may be able in reducing noise in applications with little speech degradation in diverse noise environments by reducing the residual noise and improve the intelligibility of speech. MATLAB routines are developed for performing this research.

Index terms- Time-Frequency analysis, Filter banking, Robust detection, Wavelet banking, perceptual wavelet packet transform

-----◆-----

1 INTRODUCTION

The performance of the automatic speech processing systems degrade drastically when confronted with a great adverse noise conditions such as background noise and microphone distortions. For this reason there is a strong demand for quality reduction algorithms capable of operating at very low signal to noise ratio in order to combat various forms of noise distortion. The solutions can be classified into two [1] main areas a) nonparametric; usually remove an estimate of the distortion from the noisy features, and b) statistical model based speech enhancements, statistical model based speech enhancement utilizes a parametric model of the signal generation process [2]. This paper research is based on a speech enhancement system which can be based on subtractive type algorithms. By subtracting the noise estimation from the noisy speech this system estimates the short time spectral magnitude of speech. One of the important parts of our main research (including our underway research) consists a generalized perceptual time-frequency subtraction method based on the masking properties of the human auditory system [4], this works in conjunction with a perceptual wavelet packet transform (PWPT) to reduce the effect of noise contamination. We describe in this paper is the proposed MATLAB implemented perceptual wavelet filter bank architecture. The main theme in this proposed method is the use of PWPT to approximate 24 critical bands of the human auditory system up to 16 kHz. It enables the components of complex sound to be appropriately segregated in frequency and time in order to mimic the frequency selectivity and temporal masking of the human auditory system. This proposed MATLAB implemented method uses PWPT to analyze for the application of improving the perceptual quality of the final processed speech. "Brain stem speech evoked potentials"

research data [8] was partly required for this implementation of this paper research.

2 PROPOSED PERCEPTUAL WAVELET FILTER BANK ARCHITECTURE

Architecture for the perceptual wavelet filter bank: To design this algorithm for enhancing speech a well built psychoacoustic model of the ear which has an unsurpassed capability to adapt to noise. In this a new human auditory model that adapts to the basic structure of traditional auditory model but replace the time invariant band pass filters with WPT in order to mimic the time-frequency analysis of the critical bands according to the hearing characteristics of human cochlea [10]. A PWPT is used to decompose the speech signal from 20 Hz to 16 KHz into 24 frequency sub-bands that approximate the critical bands, efficient seven level tree structures is implemented. This is given in the Fig 1. Two channel wavelet filter banks are used to split the low pass and high pass bands as opposed to only the low pass and high pass bands in the usual wavelet decomposition. Advantages: first, Smoothness property of wavelet is determined by the number of vanishing moments: more the vanishing moments the stringent bandwidth and stop band attenuation of each sub-band and can be more close approximation by using the wavelet decomposition. Second, according to the psychoacoustic study of human ears a frequency to bark transformation [5] needs to be performed which can be accomplished in audio processing systems by dividing the frequency range into critical bands. Using the perfect reconstruction filter bank with finite length filters using different wavelets for the analysis and synthesis scaling functions [5, 6, 7]. Let $H(z)$ and $G(z)$ be the low pass (LP) and high pass (HP) transfer functions, before the decimation by two operation in each stage of the analysis filter bank. $F(z)$ and $J(z)$ be the LP and HP transfer functions, after the up

sampling by two operation in each stage of the synthesis filter bank. Then the analysis and synthesis filter banks are related by

$$\begin{aligned} g(n) &= (-1)^n f(n) \leftrightarrow G(z) = F(-z) \\ j(n) &= -(-1)^n h(n) \leftrightarrow j(z) = -H(-z) \end{aligned} \quad (1)$$

The relationship between the LP and HP filters reduces the number of filters to be implemented for each stage of the two-channel filter bank by half. Once the LP filters, $H(z)$ and $F(z)$ are designed the HP filters $G(z)$ and $J(z)$ can be derived from the equation (1). According frequency selectivity related to critical band, temporal resolution of the human ear, and regularity property of wavelets debauchies wavelet basis is chosen as prototype filter and a seven stage WPT is adopted to build perceptual wavelet filter bank. Table I shows the mapping of the PWPT coefficients in each stage. Table II shows the comparison of lower (f_l) and upper (f_u) frequencies, center frequency (f_c) and bandwidth (Δf) in hertz between the critical band rate and the proposed perceptual wavelet packet tree scale. The critical band rate Z_B in bark is approximated by where the frequency, f , is measured in Hz.

$$Z_B = 13 \tan^{-1}(7.6 \times 10^{-4} f) + 3.5 \tan^{-1}(1.33 \times 10^{-4} f)^2 \quad (2)$$

In case of bandwidths of the critical bands and the perceptual wavelet packet tree, critical bands have constant width at approximately 100 hz for centre frequencies upto 500 hz, and the bandwidths increase as the centre frequency increases. The critical bandwidth (CBW) in hertz is calculated by

$$CBW(f) = 25 + 75 (1 + 1.4 \times 10^{-6} f^2)^{0.69} \quad (3)$$

Figure 2 compares the absolute threshold of hearing (ATH) [8,9] in hertz, critical band scale[6], and perceptual wavelet packet scale. The ATH characterizes the amount of energy needed in a pure tone such that it can be detected by a listener in a pure tone such that it can be detected by a listener in a noiseless environment. The table II and figures given it makes clear regarding the proposed perceptual wavelet packet tree can closely mimic the experimental critical bands. The parameters of the discrete WPT filter used to derive the plots of figures are determined based on the auditory masking properties.

3 CONCLUSION

The system can consist of functional stages working cooperately to perform perceptual time-frequency subtraction[7,9] by adapting the weights of the perceptual wavelet coefficientcents. The noisy speech is first

decomposed into critical bands by perceptual wavelet transform. We have developed MATLAB implementation. With this proposed wavelet filter banking architechure we proved that we can mimic the experimental critical bands. So this may be able in reducing noise in applications with little speech degradation in diverse noise environments by reducing the residual noise and improve the intelligibility of speech. We are under research for developing a speech enhancement system for the application of this wavlet filter bank architecture applicable to GPTFS method [8,9] for noise reduction & unvoiced speech [8] will also be enhanced by a soft thresholding scheme.

Author's biography: Mr. Ranganadh Narayanam, is currently an assistant professor in Dept of ECE in Marri Laxman Reddy Institute of Technology,Hyderabad. Mr. Narayanam had done research as research assistant in University of Texas at San Antonio, University of California at Berkeley and University of Ottawa under Dr. CJ Qian,Dr. Parimal Patel,Dr. Artyom Grigoryan, Dr. Sos Agaian, Dr. Ben Inglis and Dr. Hilmi Dajani in the areas of contorl systems,signal processing,digital systems and neurological signal processing. This research was partially guided by Dr. Hilmi Dajani of university of ottawa and Dr. Sos Agaian of University of Texas at San Antonio. Ranganadh Nnarayanam can be contacted at Marri Laxman Reddy Institute of Technology,Department of Electronics and Communications Engineering,Hyderabad,AP,India. rnara100@uottawa.ca, rnara100@gmail.com

4 REFERENCES

- [1] In-Chul Yoo and Dongsuk Yook, "Robust voice activity detection using the spectral peaks of vowel sounds". ETRI Journal,Volume 31,Number 4,August 2009.
- [2] R. Nicole, J. Sohn, N.S. Kim, and W. Sung, "A statistical Model Based Voice Activity Detection," *IEEE Signal Process. Lett.*, vol. 6, 1999, pp. 1-3.
- [3] Javier Ramirez,Jos C segura, Carmen Benitez,Angel de la torre, Antonio Rubio, "Efficient voice activity detection algorithms using long-term speech information", J. Ram_irez et al. / *Speech Communication* 42 (2004) 271-287.
- [4] I. Krekule, "zero crossing detection of the presence of evoked responses", *Electroencephalography and clinical neurophysiology*, Elsevier publishing company, Amsterdam - Printed in the netherlands.
- [5] GBrsn Eduardo Mog, Eduardo Parente kbeiro, "Zero Crossing determination by linear interpolation of sampled sinusoidal signal.
- [6] Dajani, R.H., Purcell, D., Wong, W., Kunov, H., Picton, T.W. 2005. Recording Human Evoked Potentials That Follow the Pitch Contour of a Natural Vowel. *IEEE Transactions on Biomedical Engineering* 52, 1614-1618.

- [7] Johnson, K.L., Nicol, G.T., Kraus, N. 2005. Brain Stem Response to Speech: A Biological Marker of Auditory Processing. *Ear & Hearing* 26, 424-434.
- [8] Russo, N., Nicol, T., Musacchia, G., Kraus, N. 2004. Brainstem responses to speech syllables. *Clinical Neurophysiology* 115, 2021-2030.
- [9] Galbraith GC, Arbagey PW, Branski R. Intelligible speech encoded in the human brain stem frequency-following response. *NeuroReport* 1995; 6: 2363-2367.
- [10] M.S. John, T.W. Picton, MASTER: a Windows program for recording multiple auditory steady-state response. *Computer Methods and Programs in Biomedicine* 61 (2000) 125–150, Elsevier.

Table 1 perceptual wavelet filter banks coefficients.

Sub band Z_w	l	Coefficients k_a-k_b	Transform stage j
1	1	0-0	7
2	1	1-1	7
3	1	2-2	7
4	1	3-3	7
5	1	4-4	7
6	1	5-5	7
7	1	6-6	7
8	1	7-7	7
9	2	8-9	6
10	2	10-11	6
11	2	12-13	6
12	2	14-15	6
13	2	16-17	6
14	2	18-19	6
15	4	20-23	5
16	4	24-27	5
17	4	28-31	5
18	8	32-39	4
19	8	40-47	4
20	8	48-55	4
21	8	56-63	4
22	16	64-79	3
23	16	80-95	3
24	32	96-127	2

Table 2 Critical band rate Z and perceptual wavelet filter banks W

Z	Bark Scale	Bark Scale	Bark Scale	Wavelet scale	Wavelet scale	Wavelet scale
	$[f_l f_u]$	f_c	Δ_f	$[f_l f_u]$	f_c	Δ_f
1	[0 100]	50	100	[0 125]	62.5	125
2	[100 200]	150	100	[125 250]	187.5	125
3	[200 300]	250	100	[250 375]	312.5	125
4	[300 400]	350	100	[375 500]	437.5	125
5	[400 510]	450	110	[500 625]	562.5	125

6	[510 630]	570	120	[625 750]	687.5	125
7	[630 770]	700	140	[750 875]	812.5	125
8	[770 920]	840	150	[875 1000]	937.5	125
9	[920 1080]	1000	160	[1000 1250]	1125	125
10	[1080 1270]	1170	190	[1250 1500]	1375	250
11	[1270 1480]	1370	210	[1500 1750]	1625	250
12	[1480 1720]	1600	240	[1750 2000]	1875	250
13	[1720 2000]	1850	280	[2000 2250]	2125	250
14	[2000 2320]	2150	320	[2250 2500]	2375	250
15	[2320 2700]	2500	380	[2500 3000]	2750	500
16	[2700 3150]	2900	450	[3000 3500]	3250	500
17	[3150 3700]	3400	550	[3500 4000]	3750	500
18	[3700 4400]	4000	700	[4000 5000]	4500	1000
19	[4400 5300]	4800	900	[5000 6000]	5500	1000
20	[5300 6400]	5800	1100	[6000 7000]	6500	1000
21	[6400 7700]	7000	1300	[7000 8000]	7500	1000
22	[7700 9500]	8500	1800	[8000 10000]	9000	2000
23	[9500 12000]	10500	2500	[10000 12000]	11000	2000
24	[12000 15500]	13500	3500	[12000 16000]	14000	4000

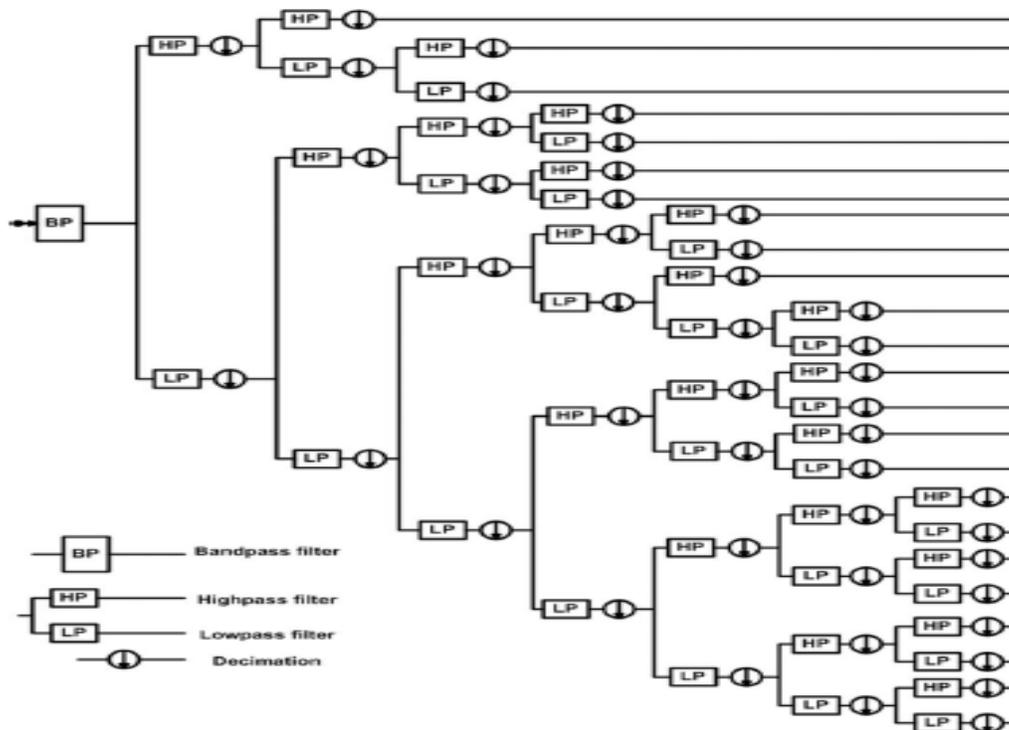


Figure 1. Perceptual wavelet packet decomposition tree (PWPT)

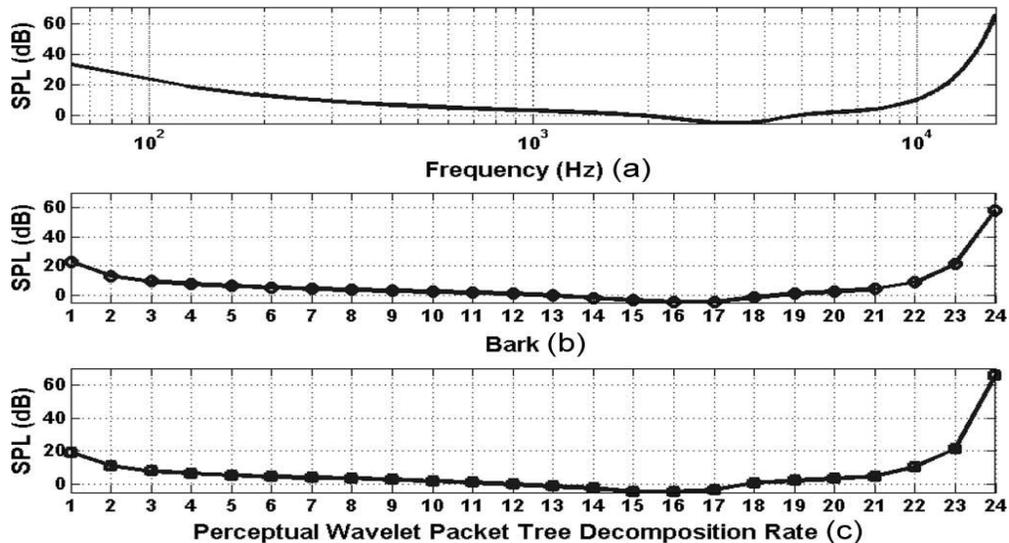


Figure 2. ATH in (a) frequency, (b) bark, and (c) perceptual wavelet packet tree scales.

Service-Oriented Channel Allocation For Maximum Reliability

S. Satish Kumar, (M.Tech)

ABSTRACT —Mobile computing involves bulk data transmission over the transmission media. To achieve highly reliable data transmission, wireless mobile networks require efficient reliable link connectivity, regardless of terminal mobility and, thus, a reliable traffic performance. Mobile networks consist of mobile hosts, base stations, links, etc. that are often vulnerable to failure. It is desirable to design a reliable network, in terms of services of both the base stations and the communication channels of the network, for the reliable transmission of the data. The objective of this study is to design an appropriate reliability-based model for channel allocation that retains the overall system reliability with acceptable system performance. The system may achieve acceptable performance not only during normal operations but also under various component failures. A genetic algorithm, which is a search procedure based on evolutionary computation, is suited to solve a class of complex optimization problems. The potential of the genetic algorithm is used, in this paper, to improve the reliability of the mobile communication system. The proposed model designs a reliable mobile communication system, irrespective of the mobile hosts that change their position due to mobility. A simulation experiment to evaluate the performance of the proposed algorithm is conducted, and results reveal the effectiveness of this model.

Index Terms: Byzantine failure, channel allocation, channel reuse, failure, genetic algorithm (GA), handoff, reliability.

◆

I. INTRODUCTION

A CELLULAR system divides a geographical communication area into smaller regions called cells, which are usually hexagonal for analytical and experimental purposes. A typical mobile network environment consists of cells, each of which is serviced by a base station (BS) located at the center of the cell. The BS provides a connection end point for the roaming mobile hosts (MHs). The BS is interconnected by wired or wireless media. The channel-allocation problem deals with the allocation of frequency channels of the given network to the MHs. Two important concepts in channel allocation are cellular reuse of channels and handoff. The fundamental and elegant concept of cells relies on the channel or frequency reuse, i.e., the usage of the same channel by different MHs separated by a minimum distance, without interfering with each other (cochannel interference). Handoff occurs when a user moves from the coverage area of one BS to the adjacent one while it is still involved in communication. A new channel will be assigned to the MHs to continue the ongoing communication. The new channel may be within the same cell (intracell handoff) or in a different cell

(intercell handoff). These issues are important in microcellular systems where the cell radius is small.

A channel-allocation algorithm consists of two phases: 1) channel acquisition and 2) channel selection. The task of the channel acquisition phase is to collect the information of

free available channels from the interference cells and ensure that the two cells within the minimum reuse distance do not share the same channel. The channel-selection phase deals with the selection of a channel from the available free channels to get better channel utilization in terms of channel reuse.

The growing importance of mobile networks has stimulated active research into how data can reliably be transmitted over the mobile communication network. This approach suggests allocating channels to the MHs in the presence of various failures in the form of uncertainties. The failure includes signal fading, channel interference, weak transmission power, path loss, etc. This paper suggests a novel idea of channel allocation based on the reliability aspect of the system.

Reliability is the ability of a system to successfully perform its functions in routine and in hostile or unexpected circumstances.

Reliability is the probability that the network, with various components, performs its intended function for a given time period when operated under normal (or stated) environmental conditions. The unreliability of a connection is the probability that the experienced outage probability for the connection is larger than a predefined maximum tolerable value. The connection reliability is related to the traffic parameters.

The MH changes its access point time to time. This instance poses several challenges in terms of ensuring system

reliability. The increasing reliance on wireless networks for information exchange makes it critical to maintain reliable communications. Even a short downtime may cause substantial data loss; thus, these networks require high level of reliability. Reliability is a crucial parameter, because any failure will not only has direct cost on maintenance but may also result in dropped calls and terminated connection. This condition may be more catastrophic in mobile computing, because it may result in Byzantine failure. Failures that inhibit communications or result in the loss of critical data are of immense importance.

The reliability-based channel-allocation model rarely figures in the literature; however, some of the models that address the other reliability issues in cellular networks have been mentioned in brief here. An optimal forward-link power allocation model for data transmission was proposed .A soft handoff/power distribution scheme had been proposed for cellular CDMA downlinks, and its effect on connection reliability had been studied by Zhao et al.. A neural network-based multicast routing algorithm was proposed by Vijay et al. to construct a reliable multicast tree that connects the participants of a multicast group. A protocol called the reliable mobile multicast protocol was proposed in to provide reliable multicast services for mobile IP networks. The mobility agent in the mobile IP was extended to assist reliable multicasting for mobile devices.

In recent years, the applications of a genetic algorithm (GA), which is a useful search procedure for optimization problems, have attracted the attention of researchers of various disciplines as a problem-solving tool. The GA is a search procedure based on the natural evolution. The GA has successfully been applied for various optimization problems for which no straightforward solution exists. Researchers of mobile computing have used the GA for channel-allocation problems. The GA has also been extensively applied for the task-scheduling problem in distributed computing systems.

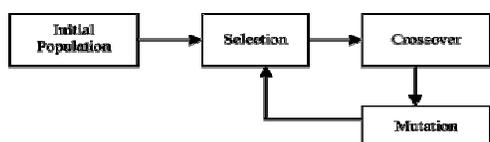


Fig. 1. Operations in the GA

II. GENETIC ALGORITHMS

The GA is a search procedure based on the principle of evolution and natural genetics. The GA combines the

exploitation of past results with the exploration of the new areas of the search space by using the “survival-of-the-fittest” technique combined with a structured yet randomized information exchange. In each new generation, a set of strings is created by using information from the previous ones. Occasionally, a new part is tried from the good measure. The GA is randomized, but it is not a simple random walk. In GAs, we start with an initial population, which is derived from the solution space. Genetic operators are then applied on the population for the appropriate mixing of exploitation and exploration. A selection strategy is used to carry forward the better population for reproduction. A simple GA consists of an initial population followed by selection, crossover, and mutation operations as shown in Fig. 1.

- 1) **Initial Population:** Initial population is the set of potential solutions to the problem. To start with, the number of solutions is generated by using any method (e.g., greedy). Borrowing the terminology from genetic engineering, the population is also called a chromosome or a string. On the initial population, various genetic operators are applied in GA.
- 2) **Selection:** The selection operation selects good results among the chromosomes by using some objective function (fitness function). The fitness function is used to rank the quality of the chromosomes. A fitness value is assigned to the chromosome, and the chromosome is evaluated with this value for its survival. The fitness of the chromosome depends on how well that chromosome solves the problem at hand. A chromosome (string) with a higher value has a higher probability of contributing to one or more offspring in the next generation.
- 3) **Crossover:** The idea of crossover is to swap part of the information between a pair of chromosomes to obtain the new chromosome. Simple crossover may proceed in two steps. First, members of the newly reproduced strings in the mating pool are mated at random. Second, each pair of strings undergoes crossing over as follows. An integer position k along the string is uniformly selected at random between 1 and the string length less than one $[1, l - 1]$. Two new strings are created by inclusively swapping all characters between positions $k + 1$ and l [3] .
- 4) **Mutation:** In mutation, a chromosome is slightly randomly altered to get a new chromosome. The mutation operator is used to introduce a new

genetic material (e.g., 0 or 1). As a result of its generality, it is an insurance policy against the premature loss of important notions. The probability of applying mutation is often very low. Mutation rates are normally small in natural populations.

A. GA-Based FTCA Model

The FTCA algorithm is designed under the resource planning model, i.e., primary channels are initially preallocated to each cell. Furthermore, the secondary (borrowed) channels must be returned to the cell from which it has been borrowed as soon as the communication is over.

Each cell has a set of reserved channels (in proportion to primary channels), which will immediately be given to a crossing over MH (to handle handoff). However, at the same time, the cell searches for a new channel. As soon as it gets the new channel, it is allocated to the crossed over MH so that there served channel pool is intact.

For experimental purposes, the MHs are randomly distributed among the cells in proportion to the number of channels per cell. It is assumed that the MH movement across the cells is stochastic.

- 1) **Encoding Used:** Each cell is represented by a chromosome. A chromosome is an array of length 14. The first location of the chromosome array represents the number of blocked hosts. The second location of the chromosome array is for the number of free channels. The next six locations contain the information about the channel lending to six neighbor cells. The last six locations contain the information about the channel borrowing from six neighbor cells. The chromosome of a cell and the chromosomes of its six neighboring cells form a matrix of 7 * 14, which is called a super chromosome. Chromosomes are combined into a super chromosome, and all the super chromosomes together give the information of the whole network. All GA operations are performed on the super chromosome.

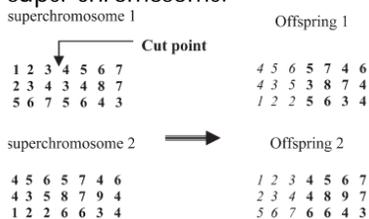


Fig 2 Crossover operation

- 2) **Crossover:** The crossover operation occurs between two superchromosomes (two matrices) to generate two offspring from them i.e., two new matrices. After this step, we get two new different chromosomes. In Fig. 2, two example superchromosomes are taken, and the crossover operation is illustrated. Crossover site is the cut point in the figure..

III. PROPOSED MODEL

In the mobile network, the system is potentially confronted with a wide range of path characteristics to each receiver e.g., different delays, link failure rates, packet losses, and competing congestion on the paths to the different receivers. Different users perceive different channel quality based on their location. The concern here with the link failure rate is in terms of the failure of the BS and channel assigned to the MH for communication.

The work proposed here considers the channel allocation based on the failure rate of the BS and the channel. With the failure of the BS, we mean the total interference level of signals received from the terminal equipment at the BS, the strength of the transmission power, the signal-to-noise ratio between the terminal equipment and the BS, etc.

Certain assumptions have been laid down in the model. As with most of the channel-allocation models, cells are assumed to be hexagonal for simplification and analytical reasons. Each cell has one BS that is responsible for allocating the channels for the hosts inside the cell and the crossing-over hosts to this cell. For experimental purposes, MHs are randomly distributed among the cells, depending on the capacity of the cell. It is assumed that the MHs' movement across the cells is stochastic. The channels are assigned to the cells according to the initial requirement of the network traffic. The probability of applying mutation is often very low. The main weakness of mutation in the channel-allocation problem is the taking-borrowing decisions ahead of time that may result in nonoptimality for two reasons: 1) Their effectiveness is not measured in the fitness function,

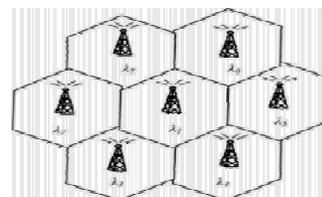


Fig 3: Bss with different failure rates

and 2) these decisions degrade the future quality of service [3], Each cell has a set of reserved channels that will immediately be given to a crossing-over MH .The performance of the algorithm is evaluated by measuring the maximum reliability value of the simulated model for the allocation.

The proposed algorithm exploits the potential of the GA to improve the reliability of the communication network system by assigning the channels to the MHs based on the reliability computation. The computation of the reliability parameter depends on two factors: 1) the reliability of the BS and 2) the reliability of the channels. The assignment of the channels to the MHs based on the reliability parameter enhances the overall reliability of the mobile network system.

A. Explanation of the Model: The reliability of the communication session depends on the services of the BSs and the links (channels) over a time T, in which the communication is made between the MHs and the corresponding node. The availability of these services depends on the failure rates of the devices (BS) and the links (channels). As previously mentioned, the failure of the BS is determined by various factors such as the total interference level of signals received from the terminal equipment at the BS, the strength of the transmission power, and the signal-to-noise ratio between the terminal equipment and the BS. The failure of the channel is determined by the traversal time of a physical path, which is its MMRT. We, in this model, have chosen the reliability parameter to be represented by exponential distribution.

The reliability of the BS over time t is $e^{-\lambda t}$, where λ is the failure rate of the BS, and t is the time of a session i.e., in which the BS is involved in communication between the terminal devices.

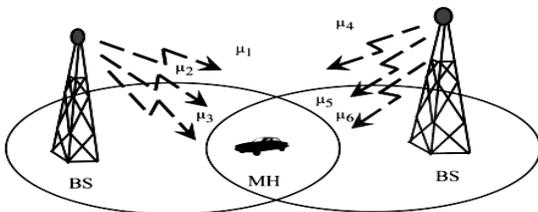


Fig.4. Network- assigning channels with different failure rates

If the number of BSs used in the network system for one whole session is m, then the reliability of all the BSs R_B in the network for the session is

$$R_B = EXP \left[- \sum_{k=1}^m \lambda_k t_k \right]. \quad (1)$$

This equation is due to the fact that the different BS with different failure rates (λ) are involved over the different time period in one session.

Similarly, if the number of total channels used in one session is n, then the reliability R_C of all these channels for that session is

$$R_C = EXP \left[- \sum_{i=1}^n \mu_i t_i \right] \quad (2)$$

where μ is the failure rate of the channel (see Fig. 4).Note that the total time taken in a session is T and is evaluated as

$$T = \sum_{i=1}^n t_i + \sum_{k=1}^m t_k. \quad (3)$$

The GA is used as a tool for optimizing (maximizing) the reliability, for both the BSs and the channels, in the proposed model. The population with better reliability value, in each generation, will participate for reproduction in one or more of the next generations. To observe the effect of communication time on the reliability of the designed network system, an experiment has been conducted for different sessions over the different time periods. An experiment is conducted for the new initiated calls and for the handoff calls.

B. Fitness Function: Based on (1) and (2), the total reliability R_T of the network system for a communication session is given by

$$R_T = R_B \times R_C. \quad (4)$$

To obtain the best reliability for the designed network system, the reliability R_T in (4) will be maximized. This function gives the total reliability of a communication session at any time T.

C. Algorithm

This section proposes a channel-allocation algorithm to optimize the reliability of the network system using the

GA. The algorithm uses a channel-allocation strategy similar to the one with reliability optimization. The algorithm is given as follows.

1. Input the total number of channels and the MHs.
2. Assign channels to each cell based on the initial demand.
3. Input generation_no. // for how many generations to carry on the experiment.
4. Initialize generation_index =0. // used as the index.
5. Initialize Max_system_reliability =0.
6. Create the initial population.
7. Allocate channels to hosts based on the strategy.
8. Repeat Steps 9–14 until generation_index= generation_number.
9. Perform the genetic operations as in Section II.
10. Score the population based on the reliability fitness function. Select the best superchromosome as the current superchromosome.
11. Output current_system_reliability resulted in the current generation.
12. Increment generation_index.
13. 14.If(current_system_reliability>Max_system_reliability)Max_system_reliability=current_system_reliability.
14. Output Max_system_reliability.

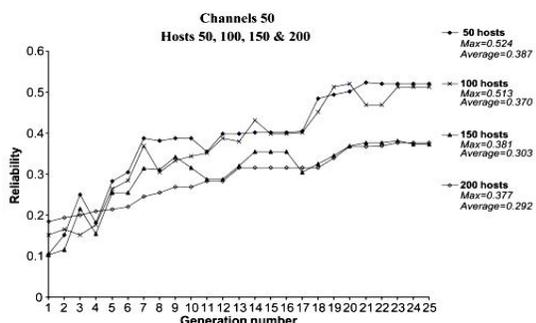


Fig.5 Fifty channels with varying numbers of hosts

IV. EXPERIMENTAL EVALUATION

In this section, the performance of the proposed algorithm is evaluated. The experiment is conducted up to 25 generations. It has been observed that the solution converges by 25 generations. The experiments have been designed by writing programs in C++.

- 1) Simulation Parameters: The simulation parameters used in the experiment are listed as follows.
 - The simulated cellular network consists of 20 cells.
 - The total number of channels and hosts in the network are varying.
 - The reserved channels, for all the experiments, are 30% of the total number of channels and are distributed among the cells in proportion to the distribution of the MHs. For example, in the experiments with 50, 100, 150, and 200 channels, the reserved channels are 15, 30, 45, and 60, respectively
 - The handoff probability is considered to be 30%, which is in conformity with that of the reserved channels. The results are represented in the performance graphs, where the x-axis represents the generations, and the y-axis denotes the reliability value.
- The experiment is conducted for random values (ranges) of BS failures λ and channel failures μ . The maximum value obtained over the generations is taken as the solution. The input values are as follows.
- $\lambda = 0.1 - 0.3$, and $\mu = 0.4 - 0.8$;
 - Number of channels: 50, 100, 150, and 200;
 - Number of hosts: 50, 100, 150, and 200.

Furthermore, we conducted the experiment by varying the values of λ and μ . First, it is conducted when $\lambda = 0.1 - 0.3$ and $\mu = 0.5 - 0.9$. Simulation has been carried out again for the four sessions, and the average value is shown in the graph in Fig. 10

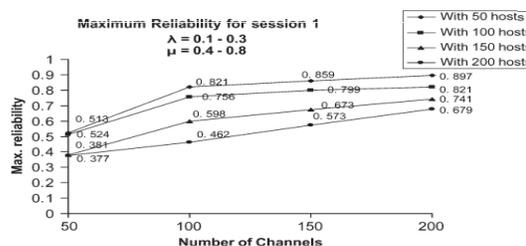


Fig6:Result of session1

V. OBSERVATIONS

Before making our concluding remarks, the following observations have been derived from the results obtained in Section IV.

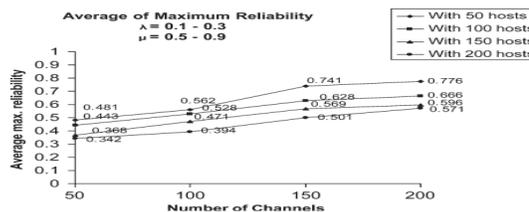


Fig7: Average results of sessions

- Both the maximum reliability value and the average reliability value increases over the generations, as shown in Figs. 7.

VI. CONCLUSION

In this paper, a reliability-based model that uses the GA to optimize the reliability in mobile computing network has been proposed. The proposed model is an effective approach to make the network connections more reliable. It has been observed that the well-managed and efficient usage of the better channels (with lower failure rates) and delivering them to the MHs greatly increases network reliability. The performance of the proposed model has been evaluated by conducting the simulation experiment. It is found that, over the generations, both maximum reliability and average reliability increase, and the result converges after certain generations. The model cannot be compared with any other method, because no other work conducts the channel allocation based on reliability values. The proposed model can be incorporated with other similar models to increase their reliability and effectiveness. In the future, we intend to observe the effect of increasing the reliability on the other quality-of-service parameters of the network system.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers, whose valuable comments and suggestions have resulted in the improved quality of this paper.

REFERENCES

- [1] C. Siva Ram Murthy and B. S. Manoj, *Ad Hoc Wireless Networks Architectures and Protocols*. Upper Saddle River, NJ: Pearson, 2004.
- [2] Q.-A. Agrawal and D. P. Zeng, *Introduction to Wireless and Mobile Systems*. London, U.K.: Thomson Learning, 2003.
- [3] L.M. O. Khanbary and D. P. Vidyarthi, "A GA-based fault-tolerant model for channel allocation in mobile computing," *IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1823–1833, May 2008.
- [4] J. Yang, Q. Jiang, D. Manivannan, and M. Singhal, "A fault-tolerant distributed channel allocation scheme for cellular networks," *IEEE Trans. Comput.*, vol. 54, no. 5, pp. 616–629, May 2005.
- [5] L. Ortigoza-Guerrero and A. H. Aghvami, *Resource Allocation in Hierarchical Cellular Systems*. Norwood, MA: Artech House, 2000.
- [6] M. Gupta and A. K. Sachan, "Distributed dynamic channel allocation algorithm for cellular mobile network," *J. Theor. Appl. Inf. Technol.*, vol. 3, no. 3, pp. 58–63, Sep. 2007.
- [7] S. S. Mahapatra, K. Roy, S. Banerjee, and D. P. Vidyarthi, "Improved genetic algorithm for channel allocation with channel borrowing in mobile computing," *IEEE Trans. Mobile Comput.*, vol. 5, no. 7, pp. 884–892, Jul. 2006.
- [8] K. Y. Lim, M. Kumar, and S. K. Das, "Message ring-based channel reallocation scheme for cellular networks," in *Proc. Int. Symp. Parallel Architectures, Algorithms, Netw.*, 1999, pp. 426–431.
- [9] D. Zhao, X. Shen, and J.W. Mark, "Soft handoff and connection reliability in cellular CDMA downlinks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 2, pp. 354–365, Feb. 2006.
- [10] W. Liao, C.-A. Ke, and J.-R. Lai, "Reliable multicast with host mobility," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 27–Dec. 1, 2000, vol. 3, pp. 1692–1696.
- [11] C. Olivena, J. B. Kim, and T. Suda, "An adaptive bandwidth reservation scheme for high-speed multimedia wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 6, pp. 858–873, Aug. 1998.
- [12] R. Jayaram, S. K. Sen, N. K. Kakani, and S. K. Das, "Call admission and control for quality-of-service (QoS) provisioning in next-generation wireless networks," *Wirel. Netw.*, vol. 6, no. 1, pp. 17–30, Feb. 2000.
- [13] D. Tipper, T. Dahlberg, H. Shin, and C. Charnsripinyo, "Providing fault tolerance in wireless access networks," *IEEE Commun. Mag.*, vol. 40, no. 1, pp. 17–30, Feb. 2002.
- [14] T. Dahlberg and J. Jung, "Survivable load sharing protocols: A simulation study," *Wirel. Netw.*, vol. 7, no. 3, pp. 283–296, May 2001.

AUTHORS BIOGRAPHY



- S. Satish Kumar has received his B.TECH from Bonam Venkatachallamayyah Engineering College (B.V.C), Odalarevu, JNTUK in 2009 and currently pursuing masters degree program in computer science and engineering in Pragati Engineering College, JNTUK, INDIA. E-mail: sskumar789@gmail.com His area of interest includes Network security.

Mathematical Model for Performance Rating in Software industry- A study using Artificial Neural Network

E.Manokaran, S.Senthilvel, S.Subhashini, Dr.R.Muruganandham, Dr.K.Ravichandran

Abstract— Multi-criteria decision making has been one of the fastest growing areas during the last decades depending on the changing's in the business sector. Decision maker(s) need a decision aid to decide alternatives and mainly excel less preferable alternatives fast. With the help of computers the decision making methods have found great acceptance in all areas of the decision making processes. Since multi-criteria decision making (MCDM) has found acceptance in areas of operation research and management science, the discipline has created several methodologies. It is difficult to find the performance of employees by considering the factors. In these paper employees performance level is analyzed and ranked by TOPSIS and SAW methods. Obtained results are validated using ANN and the results were compared Even the problem with both the objective function is complex by solving MCDM. Hence Optimization Method of approach to a problem is analyzed and further comparison is made.

Index Terms— ANN, MCDM, Performance level, SAW, TOPSIS

1 INTRODUCTION

Multi-criteria decision analysis (MCDA), sometimes called multi-criteria decision making (MCDM) [1] is a discipline aimed at supporting decision makers faced with making numerous and sometimes conflicting evaluations. MCDM consists of constructing a global preference relation for a set of alternatives evaluated using several criteria and selection of the best actions from a set of alternatives, each of which is evaluated against multiple, and often conflicting criteria. Further the objective function with both positive and negative criteria is solved through Optimization Technique which is more complex by MCDM. Therefore, the aim of this paper is to compare the various MCDM tools to decision-making problems, to determine parameter analysis compared with other methods and the mathematical approach of solving the problem with both the function.

2 TOPSIS METHOD

Technique for order performance by similarity to ideal solution (TOPSIS) [2], one of the known classical MCDM methods was first developed by Hwang and Yoon for solving MCDM problems. TOPSIS is based on the idea, that the chosen alternative should have the shortest distance from the positive ideal solution and on the other side the farthest distance of the negative ideal solution.

The TOPSIS-method will be applied to a case study, which is described in detail.

3 SAW METHOD

Simple Additive Weighting (SAW) is probably the most used and abused MCDA method. It is intuitive and easy. Simple Additive Weighting (SAW) [3] which is also known as weighted linear combination or scoring methods is a simple and most often used multi attribute decision technique. The method is based on the weighted average. An evaluation score is calculated for each alternative by multiplying the scaled value given to the alternative of that attribute with the weights of relative importance directly assigned by decision maker followed by summing of the products for all criteria. The advantage of this method is that it is a proportional linear transformation of the raw data which means that the relative order of magnitude of the standardized scores remains equal.

4 ANN METHOD

For the validation process ANN [6] is followed. The human brain provides proof of the existence of massive neural networks that can succeed at those cognitive, perceptual, and control tasks in which humans are successful. The brain is capable of computationally demanding perceptual acts (e.g. recognition of faces, speech) and control activities (e.g. body movements and body functions). The advantage of the brain is its effective use of massive parallelism, the highly parallel computing structure, and the imprecise information-processing capability. Hence the employee stress is dealing with the biological factor ANN is the best method to validate problems associated with it. Artificial neural networks (ANN) have been developed as generalizations of mathematical models of biological nervous systems.

- E.Manokaran is currently pursuing his masters degree in Industrial Engineering in Thiagarajar College of Engineering, Madurai, India. E-mail : manokarantamilan@gmail.com
- S.Subhashini is also currently pursuing his masters degree in Industrial Engineering in Thiagarajar College of Engineering, Madurai, India. E-mail : sshashini96@gmail.com
- S.Senthilvel is currently pursuing his under graduate degree in Electronics & Communication Engineering in Thiagarajar College of Engineering, Madurai, India. E-mail : senthilvelce@gmail.com
- Dr.R.Muruganandham is an Assistant Professor in the Dept. of Mechanical Engineering, Thiagarajar College of Engineering, Madurai, India. E-mail : rmmech@tce.edu

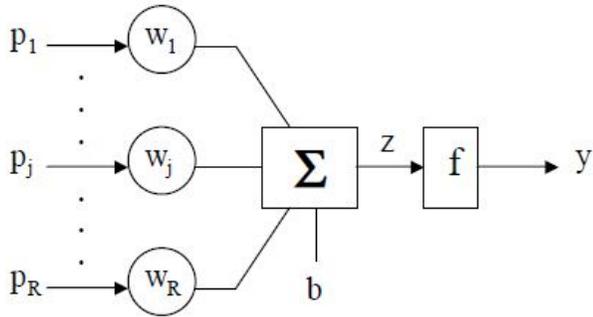


FIG 1: MCCULLOCH-PITTS MODEL OF AN ARTIFICIAL NEURON

$$Y = f (W_1.p_1 + \dots + W_j.p_j + \dots + W_R.p_R + b) \quad \dots (1)$$

$$Y = f (W.p + b) \quad \dots (2)$$

$p = (p_1, \dots, p_R)^T$ is the input column vector

$W = (W_1, \dots, W_R)$ is the weight row vector

5 DATA COLLECTION

Questionnaire was framed and detailed survey was carried among the employees and the dominant factors were only considered for the study. It is difficult to say that which employee has the best performance level from the below table.

TABLE 1

Employee No	Verbal Ability	Aptitude	Communication Skills	Academic	Data Interpretation
1	44	56	78	32	43
2	45	76	55	65	76
3	45	67	83	44	56
4	23	44	98	55	33
5	44	99	76	77	55
6	81	72	33	45	67
7	34	99	51	71	81
8	34	66	99	11	34
9	34	71	82	34	36
10	45	41	71	81	41
11	81	54	36	73	12
12	44	51	81	41	84
13	41	45	44	36	72
14	43	71	32	31	41
15	91	23	45	77	63
16	92	42	76	42	81
17	43	76	82	91	54
18	34	62	41	52	97
19	42	64	85	92	11
20	36	42	78	93	24
21	32	51	42	73	92
22	81	75	82	36	41
23	53	48	72	83	42
24	23	56	93	54	57
25	64	83	41	37	52
26	62	73	48	91	55
27	26	71	62	54	33
28	37	41	67	34	66
29	61	78	93	25	67
30	41	64	58	93	42
31	32	53	51	68	31
32	82	51	94	37	82
33	42	61	54	62	71
34	38	41	27	82	35
35	71	32	42	67	84
36	65	34	11	23	86
37	65	85	43	21	21
38	42	76	73	82	32
39	14	52	54	67	81
40	23	54	53	47	81

LIST OF EMPLOYEE'S FACTORS SCORES FOR PERFORMANCE ANALYSIS
(SINGLE OBJECTIVE FUNCTION-MAXIMIZATION)

Emp. No	Verbal Ability	Aptitude	Communication Skills	Academic	Data Interpretation	No of Days Absent (in a year)	Frequency Of Lateness
1	44	56	78	32	43	20	10
2	45	76	55	65	76	11	22
3	45	67	83	44	56	24	11
4	23	44	98	55	33	45	23
5	44	99	76	77	55	30	33
6	81	72	33	45	67	11	49
7	34	99	51	71	81	11	28
8	34	66	99	11	34	10	20
9	34	71	82	34	36	48	23
10	45	41	71	81	41	36	33
11	81	54	36	73	12	18	42
12	44	51	81	41	84	44	12
13	41	45	44	36	72	20	32
14	43	71	32	31	41	37	10
15	91	23	45	77	63	33	27
16	92	42	76	42	81	21	28
17	43	76	82	91	54	47	21
18	34	62	41	52	97	22	4
19	42	64	85	92	11	18	32
20	36	42	78	93	24	24	51
21	32	51	42	73	92	21	3
22	81	75	82	36	41	11	12
23	53	48	72	83	42	23	32
24	23	56	93	54	57	43	26
25	64	83	41	37	52	14	36
26	62	73	48	91	55	29	32
27	26	71	62	54	33	12	32
28	37	41	67	34	66	46	39
29	61	78	93	25	67	22	34
30	41	64	58	93	42	43	23
31	32	53	51	68	31	33	32
32	82	51	94	37	82	12	28
33	42	61	54	62	71	40	23
34	38	41	27	82	35	23	32
35	71	32	42	67	84	28	32
36	65	34	11	23	86	48	34
37	65	85	43	21	21	23	32
38	42	76	73	82	32	26	43
39	14	52	54	67	81	12	2
40	23	54	53	47	81	29	43

TABLE 2

LIST OF EMPLOYEE'S FACTORS SCORES FOR PERFORMANCE ANALYSIS
(TWO OBJECTIVE FUNCTION-MAX & MIN)

6 RESULTS

⌘ Above tables shows the result of employees subjected to more stressed

Comparison	Sensitivity Analysis	Large No of Criteria	Decision Maker's Support	Time Analysis
TOPSIS	3	4	2	2
SAW	2	3	3	4
ANN	3	2	3	4

7 MULTI CRITERIA DECISION METHOD-PARAMETER COMPARISON

* 1-5 for lower to best values.

Different Methods is analyzed and the grades were given based on evaluating the weight.

➤ Sensitivity Analysis:

The decision maker can make better decisions if he/she can determine how critical each criterion is. In other words, how sensitive the actual ranking of the alternatives is to changes on the current weights of the decision criteria.

➤ Large Numbers of Criteria:

As the criterion becomes large for some of the alternatives, The MCDM technique can support to its own weight age. Hence grades have been given based on the MCDM support to large number of criteria or attributes.

➤ Decision Makers Support:

According to the Decision Maker the appropriate weight-age can be given. Hence grades have been given based on the decision maker support to the MCDM.

➤ Time Analysis:

The time taken for the solving the problem with different criteria differs. Hence the grades given to MCDM tools based on their time taken to solve the problem.

8 MATHEMATICAL MODEL

In optimization models the requirements come from the relationships that must hold among the decision variables and the various static or dynamic structural elements by the nature of system operation. Each requirement leads to a constraint on the decision variables that will be expressed as a mathematical equation or inequality in the model for the problem. The model also includes any bounds (lower and/or upper) that the decision variables or some functions of them must satisfy in order to account for the physical limitations under which the system must operate.

We know that if an objective function is a cost function (profit function) we would like to minimize (maximize) it. Fortunately, it is not necessary to consider minimization and maximization

$$\left(\begin{array}{l} \text{Maximum value of } f(x) \\ \text{subject to some constraints} \end{array} \right) = - \left(\begin{array}{l} \text{Minimum value of } -f(x) \\ \text{subject to the same constraints} \end{array} \right)$$

problems separately, since any minimization problem can be transformed directly

into a maximization problem and vice versa. For example, to maximize a function $f(x)$ of decision variables x , is equivalent to minimizing $-f(x)$ subject to the same system of constraints, and both these problems have the same set of optimum solutions. Also, we can use

$$\dots (3)$$

9 CONCLUSION

It is quite clear that selection of employee's performance factor involves a large number of considerations. The use of TOPSIS method is observed to be quite capable and computationally easy to evaluate and select significant effect of stress from a given data. TOPSIS method uses the measures of the considered criteria with their relative importance in order to rank the employee with respective results. Thus, this popular MCDM [4] method can be successfully employed for solving any type of decision-making problems having any number of criteria and alternatives in the manufacturing domain. The obtained results were compared both with TOPSIS, SAW [3] and ANN [5], thus the employee with high performance level are ranked. The extension of this paper is validated by ANN. The parameter analyses for the MCDM tools are also stated with different grades. For the problem with multi-Objective function MCDM tools is not sufficient, hence mathematical model is suggested and still improvement in mathematical model can be made. As a future scope, objective function with ranges of problem with different constraint can be formulated.

REFERENCES

- [1] R. Venkata Rao, PhD, Department of Mechanical Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, Springer. "Decision Making in the Manufacturing Environment Using Graph Theory and Fuzzy Multiple Attribute Decision Making Methods".
- [2] Vijay Manikrao Athawale, Shankar Chakraborty Proceedings of "A

TOPSIS RESULTS

Rank	Employee no	Ci* value
1	5	0.7618
2	17	0.7380
3	38	0.6662
4	7	0.6516
5	29	0.6499
6	19	0.6465
7	22	0.6257
8	3	0.6191
9	24	0.6159

SAW RESULTS

Rank	Employee no	Ci* value
1	5	0.72235
2	17	0.711171
3	32	0.6968
4	26	0.688155
5	16	0.663965
6	29	0.662241
7	2	0.645196
8	22	0.62688
9	38	0.627662

ANN RESULTS

Rank	Employee no	Ci* value
1	5	862065042
2	17	809685214
3	32	733352226
4	26	668567137
5	16	614285375
6	7	607033713
7	2	571348249
8	29	455729722
9	22	452097024

"TOPSIS Method-based Approach to Machine Tool Selection" the 2010 International Conference on Industrial Engineering and Operations Management Dhaka, Bangladesh, January 9 – 10, 2010.

- [3] Alireza Afshari, Majid Mojahed and Rosnah Mohd Yusuff. "Simple Additive Weighting approach to Personnel Selection problem" International Journal of Innovation, Management and Technology, Vol. 1, No. 5, December 2010 ISSN: 2010-0248.
- [4] Farnaz Akhavi and Caroline Hayes "A Comparison of Two Multi-Criteria Decision-making Technique", IEEE, 2003, pp956-961.
- [5] Coryn A.L. Bailer-Jones Max-Planck-Institute for Astronomie, Konigstuhl "An introduction to artificial neural networks", pp.69-117.
- [6] Ajith Abraham Oklahoma, The Book "Artificial Neural Networks" State University, USA.
- [7] S. R. Gangurdel, M. M.Akarte, "Ranking of Product Alternatives Based on Customer-designer Preferences" Proceedings of the 2010 IEEE IEEM978-1-4244-8503-1110, IEEE.
- [8] Anil Jadhav, Rajendra Sonar "Analytic Hierarchy Process (AHP), Weighted Scoring Method (WSM), and Hybrid Knowledge Based System (HKBS) for Software Selection: A Comparative Study" Second International Conference on Emerging Trends in Engineering and Technology, ICETET-09, Indian Institute of Technology Bombay, India.
- [9] Jun-Jie Lv, Yuan-Zhuo Wang "A Ranking Method for Information Security Risk Management based on AHP and PROMETHEE", This paper is supported by National Natural Science Foundation of China (No.60803123, No. 60873193). 978-1-4244-5326-9/10©2010 IEEE.

Wireless Systems “Advanced 4G Technology”

D.HariPrashanth

Abstract- The recent revolution created by 3G technology paved the world to enter into a new fastest communication and fully flexible environment of internet at high speed. Now its time for even advanced technology for future generations called as the 4G technology. The components of 4G Technology, advanced materials used in it, mode of communication, upgraded versions of previously used components in 3G, and also the advantages of the 4G systems were about to be discussed.

Index Terms- 1) Introduction 2) Overview of 3G 3) Components used in 3G 4) Evolution Of 4G 5) Advantages of 4G 6) Components Of 4G 7) OFDMA 8)MIMO 9) Smart Antennas 10)SDR(Software Defined Ratio) 11) IP v6.0 12) Spectral efficiency 13) Working Of 4G 14) Conclusion.

1. Introduction:

The recent advancements in the communication technologies with special reference to the wireless communication technology has paved the way for faster, more reliable modes of data transfer and communication means. And now it's time for the deployment of even advanced wireless communication system called as **4G (4th generation)** technology which was yet to emerge within months/years.

4G technology offers high rate of data transfer at low cost than in 3G and also accessing applications with a high degree of customization and personalisation of user applications.

2. Overview of 3G:

3G or 3rd generation mobile telecommunications is a generation of standards for mobile phones and mobile telecommunication services fulfilling the **International Mobile Telecommunications-2000 (IMT-2000)** specifications by the International Telecommunication Union. Application services include wide-area wireless voice telephone, mobile

Internet access, video calls and mobile TV, all in a mobile environment.

Several telecommunications companies market wireless mobile Internet services as 3G, indicating that the advertised service is provided over a 3G wireless network. Services advertised as 3G are required to meet IMT-2000 technical standards, including standards for reliability and speed (data transfer rates). To meet the IMT-2000 standards, a system is required to provide peak data rates of at least 200 Kbit (about 0.2 Mbit/s). However, many services advertised as 3G provide higher speed than the minimum technical requirements for a 3G service. Recent 3G releases, often denoted 3.5G and 3.75G, also provide mobile broadband access of several Mbit/s to smart phones and mobile modems in laptop computers.

3. Components of 3G:

- the UMTS system, first offered in 2001, standardized by 3GPP, used primarily in Europe, Japan, China (however with a different radio interface) and other regions predominated by GSM 2G system infrastructure. The cell phones are typically UMTS and GSM hybrids. Several radio interfaces are offered, sharing the same infrastructure:

- The original and most widespread radio interface is called W-CDMA.
- The TD-SCDMA radio interface was commercialised in 2009 and is only offered in China.
- The latest UMTS release, HSPA+, can provide peak data rates up to 56 Mbit/s in the downlink in theory (28 Mbit/s in existing services) and 22 Mbit/s in the uplink.
- the CDMA2000 system, first offered in 2002, standardized by 3GPP2, used especially in North America and South Korea, sharing infrastructure with the IS-95 2G standard. The cell phones are typically CDMA2000 and IS-95 hybrids. The latest release EVDO Rev B offers peak rates of 14.7 Mbit/s downstream.

The above systems and radio interfaces are based on kindred spread spectrum radio transmission technology. While the GSM EDGE standard ("2.9G"), DECT cordless phones and Mobile WiMAX standards formally also fulfil the IMT-2000 requirements and are approved as 3G standards by ITU, these are typically not branded 3G, and are based on completely different technologies.

A new generation of cellular standards has appeared approximately every tenth year since 1G systems were introduced in 1981/1982. Each generation is characterized by new frequency bands, higher data rates and non backwards compatible transmission technology. The first release of the 3GPP Long Term Evolution (LTE) standard does not completely fulfill the ITU 4G requirements called IMT-Advanced. First release LTE is not backwards compatible with 3G, but is a pre-4G or 3.9G technology, however sometimes branded "4G" by the service providers. Its evolution LTE Advanced is a 4G technology. WiMAX is another technology verging on or marketed as 4G.

4. Evolution of 4G:

In order to make smooth transition from 3G to 4G the mobile communication companies are promoting Super 3G/LTE. The companies are upgrading 3G Technology by initializing the introduction of High Speed Downlink Packet Access (HSDPA) service, which increases the downlink data rate of packet services, and by finalizing specifications for High Speed Uplink Packet Access (HSUPA), which enhances uplink speed. HSDPA and HSUPA cover area by 3-4 times relative to W-CDMA and by providing the high transmission rate with low cost per bit transmission. The main objective of the Super 3G is to construct simple, low cost system by removing the complexity from wireless network and mobile handsets. The 3G provides packet and voice services separately whereas Super 3G is based on ALL-IP network covering both packet and voice services. As from diagram we can infer that by the 2010 we would be able to achieve the 1 Gbps in motion at low speed and 100 Mbps at high speed. On December 25, 2006, NTT DOCOMO became the first in the world to achieve a packet signal speed of 5 Gbps in an outdoor test in a low-speed environment (10 km/h). The test was undertaken to demonstrate the expected maximum transmission speed in an actual cell environment, taking into account interference from peripheral cells.

5. Components of 4G:

There are some components which makes the successful 4G systems they are:

- a) OFDMA
- b) MIMO
- c) IPv6.0
- d) Spectral efficiency of 4G
- e) SDR(Software Defined Ratio)
- f) Smart antennas

6. OFDMA(Orthogonal Frequency Demux):

It captures entire energy because of capability to absorb high no. of OFDM signal subcarriers. In OFDM, as long as guard interval is long enough, all inter-symbol-interference is absorbed

And Multipath self-interference does not affect OFDM, only a few tones are affected or lost in OFDM while compared to CDMA in 3G. Implementation of equalization, interference cancellation, and adaptive antenna array algorithms is simpler in OFDM.

7. MIMO(Multi Input Multi Output):

To improve the communication performance between sender and receiver, the multiple antennas are used at both transmitter and receiver end. The signal transmitted by m antennas and signal received by n antennas and the processing of the received signal may produce significant performance improvement such as **range, quality of received signal** and **spectrum efficiency**.

8. Smart Antennas:

There are two types of smart antennas which are switched beam smart antennas and adaptive array smart antennas. Switched beam systems have several available fixed beam patterns which help in making decisions as to which beam to access at any given point of time based on the requirements of the system. While adaptive arrays allow the antenna to steer the beam to any direction of interest while simultaneously nulling interfering signals.

9. SDR(Software Defined Ratio):

A basic SDR produces a [radio](#) that is capable of receiving and transmitting a different form of radio protocol (sometimes referred to as a waveform) as per the needs just by **running different software**. A SDR will allow increasing network capacity at specific time.

10. IPv6.0:

4G wireless technology will be using mobile IPv6 which allows assigning more number of addresses than IPv4. In IPv6 each device have **own IP address**. User can keep their IP address even if user changes the access point..

11. Spectral Efficiency in 4G:

The 4G wireless technology bandwidth efficiency will be measured in terms of spectral efficiency.

Spectrum efficiency describes that the amount of information that can be transmitted over a given bandwidth in a specific communication system. It is a measure of how efficiently a limited frequency spectrum is utilized by the physical layer protocol, and sometimes by the media access control (the channel access protocol). Clearly the bit rate should be associated with an amount of spectrum. For mobile use, a good target is a network performance of 5 bit/s/Hz, rising to 8 bit/s/Hz in nomadic use.

12. Brief Note on Working Of 4G:

- a. Based on an all-IP packet switched network.
- b. Peak data rates of up to approximately 100 Mbit/s for high mobility such as mobile access and up to approximately 1 Gbit/s for low mobility such as nomadic/local wireless access, according to the ITU requirements.
- c. Dynamically share and use the network resources to support more simultaneous users per cell.
- d. Scalable channel bandwidth 5–20 MHz, optionally up to 40 MHz
- e. Peak link spectral efficiency of 15 bit/s/Hz in the downlink, and 6.75 bit/s/Hz in the uplink (meaning that 1 Gbit/s in the downlink should be possible over less than 67 MHz bandwidth).
- f. System spectral efficiency of up to 3 bit/s/Hz/cell in the downlink and 2.25 bit/s/Hz/cell for indoor usage.
- g. Smooth handovers across heterogeneous networks.
- h. Ability to offer high quality of service for next generation multimedia support.

13. Advantages of 4G:

4G technology offers high rate of data transfer at low cost than in 3G and also accessing applications with a high degree of customization and personalisation of user applications,

The user will be able to receive HD streaming of video, and the data range of 4G will be 100M/bits and 1G/bits. Broadband applications may be like wireless broadband access, Multimedia Messaging Service (MMS), video chat, mobile TV, HDTV content, Digital Video Broadcasting (DVB) demands high data rate

and the quality of service(QoS) but this type of data rate and QoS are not available in 3G technology.

The main objective of 4G technology is going to be based on OFDMA (Orthogonal Frequency Division Multiple access) modulations with MIMO (Multiple Input Multiple Outputs) and other smart antennae enhancements.

14. Conclusion:

4G technologies was still at research stages in many of the countries however companies like NTT Do Como, Mobile and Nortel Networks, and Nokia Siemens Networks successfully demonstrated the working of 4G. This technology will be deployed in the world market soon.

Details about Author:

**D.HariPrashanth,
M.S.SoftwareEngg,
VIT UNIVERSITY, Vellore.
INDIA.**

Residence address:

**105,RoyanKuttai Street,
Big Kanchipuram,
Tamil nadu-631502.**

Email: hprashanth610@gmail.com

Ph: +91-9789640904.

An expert system for Seismic data interpretation using visual and analytical tools

Neelu Jyothi Ahuja* and Parag Diwan

Abstract--Seismic images are extensively used by petroleum geologists to delineate subsurface geological structures. This interpretation is dependent on interpreter's individual skills, and expertise causing a complex seismic image to be interpreted differently by different seismologists. To overcome this uncertainty a prototype expert system SeisExpert, has been developed in which the experience and heuristic knowledge of the experts have been incorporated in the form of a set of rules. Over 120 rules have been built into a commercially available expert system shell, 'Flex' (from Logic Programming Associates, UK). Information received from the user is chained through these rules to interpret the seismic images. Additionally, analytical procedures have been provided to refine the interpretation in case of complex subsurface structures. These include statistical cross-correlation across the seismic traces and secondary attributes such as instantaneous phase and reflection strength. Using this system, several horizons were successfully tracked across seismic traces in four different seismic images containing field data from existing oil fields. Discontinuities such as faults, wherever present, were also correctly identified.

Keywords: Expert System, Seismic Data Interpretation, Rule-Base, Knowledge-Base, Seismic Attributes, Expert System Shell, Flex

1 INTRODUCTION

1.1 Petroleum Exploration

In an early twentieth century, the exploration for petroleum was carried out by petroleum geologists, whose prime concern was to locate geological structures suitable for hydrocarbon accumulation. Commercially valuable accumulations of hydrocarbons are usually found at depths of at least a few thousand meters below the ground surface. Although exact knowledge of geology at such depths can come only from drilling a bore hole, such drilling is very expensive. The deeper the structures, less is the resolution that gravity or magnetic surveys can give. Under such conditions, seismic reflection methods become useful, and play a prominent role in the search for suitable geological structures.

1.1.1 Seismic Survey: Two dimensional (2D) seismic survey is generally carried out in virgin/new areas to delineate and map the structures on a regional scale and propose locations for 3D and 4D surveys and also for exploratory drilling. An area of operation is covered with seismic profiles at regular intervals both in dip and strike directions depending upon

the orientation of the sub-surface structure inferred from the surface geological mapping. The 2D seismic survey provides an image of a geological cross-section only whereas 3D and time-lapsed 3D (4D) surveys are carried out to fill-in more details of the geological structures with greater precision. However, 2D surveys provide the basic data and hence are of utmost importance. The present study concerns only with 2D seismic data interpretation.

A typical seismic survey involves creating seismic waves by explosion in a hole in the ground. These waves travel in all directions and get reflected from certain subsurface formations and are received by a number of geophones located on the earth's surface in the vicinity of the explosion. This process is called Seismic Data Acquisition, in which amplitude of the reflected wave of a fixed frequency is recorded as a function of time which is essentially a two-way travel time from explosion source to the reflecting horizon and back to the geophones.

The data recorded at the geophones are amplitudes of reflected waves at sampling time instants (usually 2 or 4 milliseconds). These data undergo extensive processing with a view to improve signal-to-noise ratio and resolution. The processed data are finally presented in a universally accepted SEG-Y^{†1} format for interpretation. The next most significant step is the interpretation of the processed seismic

• Neelu Jyothi Ahuja is an Assistant Professor at Computer Sciences Department at University of Petroleum and Energy Studies, Dehradun, India
• Ph: 09411384390 email: neelu@ddn.upes.ac.in; neelu_j_ahuja@yahoo.com
• Parag Diwan is founder Vice Chancellor of University of Petroleum and Energy Studies, Dehradun, India, vc@upes.ac.in

[†] Glossary of terms in Appendix A

data. This involves obtaining accurate and meaningful inference from reflectors[†] which can be tied to subsurface geology in terms of structural lithology of possible hydrocarbon bearing elements. Some of the geological structures encountered are folds, faults, salt domes, pinch-outs, stratigraphic traps etc. The problem of seismic data interpretation is to track horizons reliably and delineate geological structures correctly.

There are recognized human experts who routinely interpret seismic images manually, using their expertise. However, in this process of interpretation, there is fair amount of uncertainty in terms of geological structure. A given set of seismic images are likely to be interpreted in two different ways by two different seismologists with widely varying conclusions, and these often deviate from the actual structures found. This uncertainty is partly because of extreme geological complexities and partly because there are no formal rules, and each expert uses his individualistic knowledge-base of unwritten thumb rules, that he or she has developed over the years.

The development of computer technology, with increased sophistication in recording, processing and displaying techniques, created a perfect environment for the growth of seismic technology. The availability of multitude of color coding and printing facilities further added to increased clarity in the interpretation process. Digital displays are preferred because of their superior color rendering and depth effect. But paper displays are still used for practically understanding the seismic section and interpreting it [1].

1.2 Artificial Intelligence

AI can handle complex, non-numeric problem-solving tasks where there are uncertainties or when the available information is incomplete. Its application in petroleum engineering is gaining currency and has the potential to dominate other analytical tools used in the Exploration and Production industry. The uncertainty in interpretation of seismic data has motivated us to apply artificial intelligence to this significant part of petroleum exploration. The present study is aimed at developing an expert system for seismic data interpretation. This will minimize dependence on human experts.

A WinProlog based expert system shell 'FLEX' (Logic Programming Associates, UK) has been used to develop the expert system that has been christened as SeisExpert, in which individualistic interpretation knowledge, solicited from experts, has been incorporated to form its knowledge-base. The formulated rules mimic expert's way of thinking, reasoning and decision making to generate an acceptable inference about the geological structures in a particular region. The input to SeisExpert is the seismic image and amplitude of the reflected wave as a function of two-way travel-time. The output provides information about the geological structures present in that area.

2 LITERATURE REVIEW

Analytical tools were developed for interpretation once computer processed data became available. The most elementary but very powerful technique was statistical cross-correlation of amplitude from one trace[†] with the adjoining trace. This allowed the user to move across the traces but along the same reflecting horizon[†] [2]. What appears as reflecting horizons on seismic maps could be confirmed with this technique. This allowed tracking all significant horizons in an area with high level of confidence as long as there were no discontinuities. Similar results were obtained using auto-trackers based on image-processing tools [3,4].

A seismic attribute is a quantitative measure of a seismic characteristic of interest. These attributes are calculated from the primary attribute which is the amplitude of the reflected wave. This amplitude can be viewed as the real component of a complex trace which can be uniquely calculated under usual conditions. The complex trace permits the calculation of instantaneous phase and reflection strength of the seismic signal at any instant of time. These terms will be defined later in the 'Expert System Architecture' section. These and other quantities can be displayed in a color-encoded manner which help an interpreter see their interrelationship and spatial changes [5]. Additionally, the instantaneous phase and reflection strength could also be used to track the horizons even across faults. There are a number of other attributes that have been identified but not all are useful [6-8].

Semi-automatic fault interpretation was introduced by Landmark Graphics Corporation in 1997 [9] with 'seeding' technique which allowed 'seeding' one or more fault segments on a vertical seismic section, and the automatic operation would perform a cross-correlation on a series of slanted traces derived parallel to the seeded fault segment [10]. A 'seedless' approach to fault segment extraction was presented by Bemmell and Pepper [11], where the gaps and sharp gradients from a horizon interpretation were subjected to a connected body analysis followed by feature testing to deduce likely fault candidates. Aurnhammer and Tonnie [12] described a model-based approach which reduces uncertainties in horizon correlation across faults by introducing global features based on geological constraints.

In seismic data interpretation, horizon picking is important for structural analysis, feature recognition and site appraisal. These were commonly done by hand, and are error-prone and time-consuming. A new method was devised by Harrigan et al., [13] which combined the traditional approach with a new technique using a trained artificial neural network (ANN). This method is more robust and facilitates tracking through conventionally difficult regions containing faulting and other geophysical anomalies.

Some researchers in the field claim that inherent seismic information is lost in the attribute extraction process and advocate the use of raw data instead. Benbernou and

Warwik [14] investigated the performance of ANN using both characterization methods (seismic attributes and raw amplitude data), and demonstrated how the complementarity of both can be exploited in conjunction with other geological information in a fuzzy inference system (FIS) to achieve an enhanced auto-tracking performance. Li [15] described a new classification technique to recognize and predict reservoirs from seismic data using Support Vector Machine (SVM) pattern recognition. It was demonstrated that this method is less subject to overtraining difficulties unlike ANN and can be used to distinguish between oil and gas reservoirs.

Computer assisted geo-interpretation explores the logic of human interpreters and applies it in designing expert systems. Expert Systems are early commercial successes of Artificial Intelligence. Main objective of these systems is to gather expert knowledge, represent it in appropriate format and use it like a human expert to take decisions. The system is also made capable of providing justifications and explanations for the decisions taken [16]. A system named "PROSPECTOR" [17] was developed to provide consultation services for mineral exploration. Cairn's Petrophysical Handbook [18] has described, in detail, several systems including PROSPECTOR, FACIOLOG, MUDMAN etc. A well-log interpretation system was developed at this time using artificial intelligence techniques to show what a computerized stratigraphic interpretation system can do [19]. Pitas and Venetsanopoulos [20] developed an automated knowledge-based system AGIS (Automated Geophysical Interpretation of Seismic Images), for geophysical interpretation of seismic data that can recognize various patterns working interactively with the interpreter for improved performance.

A new conceptual approach was presented by Whitney to show use of expert systems to assist in decision making process [21]. A second era of expert systems began with the developments in the field of health and medicine [22].

The user's confidence in the derived conclusion can be significantly increased by revealing internal rules that led to it. [23]. There are a few formats that have been popular for knowledge representation. Rattanaprateep and Chittayasothorn [24] presented design and implementation of a frame-based object-relational database with a tight coupling between the expert system and the external knowledge-base.

Expert systems have been developed with application in diverse fields [25-27]. Some examples are: tourist advisor system [28], as education tool [29-30], in fault diagnostics and control of power system equipment [31], testing and equipment trouble-shooting [32,33], and disaster management [34].

Many expert systems were developed in subsequent years for variety of geological and mineral exploration tasks such as Dipmeter Advisor [35,36], Laser Drilling System Optimizer [37-40].

Traditional expert systems were constructed using a single monolithic software program for a specific application which required coding the structural framework every time a system was developed. Bache et al., [41] discussed construction of an expert system shell which provided a framework for constructing application specific systems. These workers developed, an automated and interactive tool called Intelligent Monitoring System, to detect and locate seismic events. It is programmed to provide the knowledge representation framework and inference mechanisms for complex and knowledge-rich rule-based reasoning. Pan and coworkers [42] presented a general network with fuzzy logic in a large scale expert system shell which was particularly suited to deal with uncertainty in information. A fuzzy expert system called Smart-Drill, was developed to solve lost circulation problems [43], another one was for screening wells for drilling and preliminary drilling fluid selection based on different well-bore and reservoir conditions [44]. Kumar and coworkers [45] developed a shell-based expert system for the design of airborne equipment.

McCormach [46] combined ANN modeling for pattern recognition, with expert system to increase its interpretation capability. Quah and Tan [47] presented architecture of a hybrid neural network expert system shell aimed at preserving semantic structure of the system rules whilst incorporating learning capability of neural networks into the inference mechanism.

'Flex' (from Logic Programming Associates, UK) is a versatile expert system shell, which offers an open-ended knowledge-based solution to business problems. It is implemented in Prolog, which is a high-level logic-based language suited for rule-based decision making. It employs a natural language style approach to define knowledge through the provision of a dedicated Knowledge Specification Language, KSL [48].

As discussed in this section, we can see that expert systems and even rule-based expert systems have been developed in several areas but little has appeared in published literature on seismic data interpretation. Although auto horizon picking programs, fault recognition programs and several other pieces of software have been developed to deal with specific problems, there appears to be no attempt in the recent past to develop an overall system to interpret the entire seismic image. This served as motivation to undertake the present study to develop a rule-based expert system using the shell 'Flex'.

3 EXPERT SYSTEM ARCHITECTURE

The system has been designed with the following components:

1. Front-end/ User Interface
2. Expert System Shell
3. Intelligence Server

4. Analytical Application Programs

The following section presents in detail, the design and working of each of these components and elaborates their role in the operation of the system.

3.1 Front-end

The front-end serves as a user interface for the rest of the system. The product presents itself in the form of a user-friendly package that facilitates rule-based interpretation of 2D seismic sections. It displays the seismic section and queries the user to gather responses. It also invokes programs that apply analytical techniques on seismic data. The product assimilates results obtained from analytical tools to further reconfirm and fine tune the interpretation. Finally, the interpretation of the given horizon or the seismic section is displayed to the user. This part of the system has been developed in Visual Basic 6.0.

The interpretation process of the expert system can be put forth in the form of two main components of the system. These are: manual interpretation and the analytical interpretation. The manual interpretation requires only the seismic image as input and conclusions are based on the visual observations alone. On the other hand, analytical interpretation additionally requires the amplitude of the reflected seismic wave as a function of two-way-time as input, and makes use of some calculated seismic attributes to support interpretation. More details of both follow.

3.1.1 Manual interpretation

Manual Interpretation works on the principle of visual inspection of the image of a seismic section, coupled with questioning by the system. The reflecting horizons seen on a reasonably clear seismic map are in many ways comparable to the subsurface geology of the region. Therefore, close observation of the reflectors can reveal the presence of reasonable number of features of interest on the seismic map. The user can call any seismic section from the database for interpretation in the form of an image on the screen. The user can navigate between the sections available in the database and choose the desired section for display in the image box on the screen. The system then begins with the questioning process where the user is presented with questions and the responses given by him are recorded. Based on the responses, more questions are put forward. The expert system provides clues and explanations, as and when desired, to assist the user in answering questions and also provides justifications for the line of reasoning followed and conclusions obtained. The questions are presented in three

major styles: edit box[†], list box[†] and multi-box[†]. In all of these questioning styles, a button labeled 'Explain' is available with each question. This button when clicked provides the rationale behind the question being asked to the user. In the form of responses to the questions, the information is gathered and provided to rules which fire in the background. In the present case, the rules have been chained in a data-driven or forward chaining manner. In forward-chaining strategy, the system continues to collect data at each step, and builds the final results. After the question bank is exhausted and all relevant rules have fired, the expert system builds an interpretation from gathered information and displays it to the user.

3.1.2 Analytical Interpretation

Analytical interpretation process for seismic sections is invoked only when the user chooses 'seismic images with data' option. The amplitude vs. two-way travel time (TWT) data in the form of an 'excel sheet' is loaded and displayed to the user.

For analytical interpretation, certain seismic attributes are computed and used in a procedure to track a reflecting horizon across seismic traces. The most important attribute is the measured amplitude of the reflected wave which forms the basis for computation of other desired attributes such as 'instantaneous phase' and 'reflection strength'. The measured amplitude of the reflected wave depends on the nature of interface between two adjacent rock types. A statistical technique called 'Cross correlation' employs this attribute to track a reflection horizon across the seismic traces and clearly shows, slopes, slope reversals etc in the subsurface formations. This technique has been described later in the 'Implementation Notes' (Appendix B).

In order to compute the secondary attributes, the seismic trace is considered as a complex quantity with the amplitude of the reflected wave forming the real component. The imaginary component can be found, by taking Hilbert transform of the real part [5]. Some of the attributes that can be computed through this complex trace analysis are: instantaneous phase, reflection strength (envelope), apparent polarity, instantaneous frequency, weighted average frequency etc [7].

The real seismic trace $f(t)$ can be expressed in terms of time dependent amplitude $A(t)$ and phase $\theta(t)$ as

$$f(t) = A(t) \cos \theta(t) \quad (1)$$

The imaginary part, $f^*(t)$ can be expressed as

$$f^*(t) = A(t) \sin \theta(t) \quad (2)$$

Combining equations (1) and (2), the complex $F(t)$ is represented as:

$$\mathbf{F}(t) = \mathbf{f}(t) + \mathbf{f}^*(t) = \mathbf{A}(t) e^{j\theta(t)} \quad (3)$$

Instantaneous phase $\theta(t)$ can be calculated as:

$$\theta(t) = \tan^{-1} [\mathbf{f}^*(t)/\mathbf{f}(t)] \quad (4)$$

The term phase (or phase angle or phase shift) used in control literature is a function of frequency but independent of time. On the other hand, the instantaneous phase is defined for a fixed frequency but varies as function of time which is being used here. Instantaneous phase is effective in demarcating discontinuities, faults, pinch-outs and angularities clearly.

Reflection strength $A(t)$ can be calculated as:

$$\mathbf{A}(t) = [\mathbf{f}^2(t) + \mathbf{f}^{*2}(t)]^{1/2} = |\mathbf{F}(t)| \quad (5)$$

The reflection strength, also called the "amplitude of the envelope", can be calculated at each time instant in a similar manner as instantaneous phase, but is independent of it. Maximum reflection strength need not coincide with maximum phase or amplitude of the largest real trace peak. Major lithological changes between adjacent rock layers usually result in high reflection strength which facilitates tracking reflecting horizons, discontinuities etc.

3.2 Expert System Shell

The expert system shell used in the current context is available with the framework comprising of all essential components of a typical expert system namely, the knowledge-base, inference engine and working memory. Flex is an expressive and powerful expert system toolkit which supports frame-based reasoning with inheritance, rule-based programming and data-driven procedures fully integrated within a logic programming environment, and contains its own English-like Knowledge Specification Language (KSL). Sample coding has been included in the "implementation notes" (Appendix B).

3.3 Intelligence Server²

The Intelligence Server acts like a glue connecting the front-end Visual Basic interface with the expert system shell and facilitates communication between these two components. The files *edtbox.frm*, *lstbox.frm* and *mltbox.frm* are to be specifically added into the front-end to facilitate interconnection. The questions, which have been written in 'Flex' are displayed through Intelligence Server and the gathered responses follow the same route to 'Flex'.

² Proprietary component of Logic Programming Associates, UK

3.4 Analytical Application Programs

This component constitutes analytical programs that use amplitude data for seismic interpretation. The starting point of the analysis is the SEG-Y file. This file can be viewed using packages like 'Sei-See' (from DMNG Ltd). The file header holds various details such as location of the survey, sampling rate, distance between geophones, date and time etc.

Using the amplitude data, cross-correlation is carried out as detailed in "implementation notes" (Appendix B). The output in the form of a plot of the specific horizon traced is displayed on the screen which is then compared with the same horizon on the seismic image by superimposition or juxtapositioning. If there is a good match it confirms earlier manual interpretation and no further study is required. This is accomplished through a question-answer session and gathered responses are used to build the interpretation in the system. However, if only first part of the horizon matches with the later part deviating significantly, further investigation is required, as this may be due to an anomaly in the geological formation such as fault. Other attributes mentioned above have been used to confirm the presence of a fault.

As discussed earlier, the amplitude data are transformed into complexes containing both real and imaginary components (using Hilbert function, MATLAB, from Mathworks Inc.). Equations (4) and (5) were used to compute instantaneous phase and reflection strength. The resultant plots are displayed and compared as in case of cross-correlation. The findings are passed on to the system to further refine interpretation. All the above algorithms have been developed in C++ language.

The cross correlation technique is particularly useful in tracking continuous horizons, which may be inclined or flat, or folded. However, cross-correlation does not always give desired results in identifying discontinuities such as faults. This shortcoming can be overcome by using instantaneous phase and reflection strength which are characteristic of the reflecting horizon and therefore, capable of tracking horizons even across faults.

4 SYSTEM EVALUATION

The expert system described in the previous section was tested extensively using real seismic data collected from existing oil fields. The results of evaluation are discussed in this section. All geological sections selected for this study had already been interpreted and confirmed for the structural details. The different geological sections were examined using SeisExpert. Each horizon was studied by visual inspection and information was input through the question-answer sessions. Subsequently analytical tools were applied to confirm or refine the manual interpretation.

4.1 Horizon tracking using cross-correlation technique:

The process started with picking up a marker, from the first trace of the reflected wave, in this case, maximum amplitude value in that trace, along with its corresponding time. Then a time window, of 5 sampling instants was selected, with marker in the center of the window. This was cross-correlated with the adjoining trace, as demonstrated by an example described below.

Example 1: Cross-Correlation

Sample data given in the table below shows amplitudes for first two traces with marker at 2520 ms in Trace#1. The first window comprises of 5 rows with marker at the center. For the ease of handling, the original data was scaled by multiplying with 10e-3 and truncated to 2 decimal places.

Time (in ms)	Amplitude value	
	Trace #1	Trace #2
2512	-9.19	-9.16
2514	-7.20	-6.68
2516	2.56	3.39
2518	16.17	16.85
2520	25.04	24.87
2522	21.84	20.24
2524	6.49	3.73
2526	-12.73	-15.25
2528	-24.78	-25.23

Labeling data in the columns under Trace#1 and Trace#2 as x_i and y_i respectively and substituting the values in the equations, given in Appendix B, we obtain the 5 cross-correlation coefficients, a1 to a5 as follows:

- (1). 274.02 (2). 229.42 (3). 99.93 (4). 204.95 (5). 94.43

Note that the first window of five sampling instants (x_1 to x_5 and y_1 to y_5) corresponds to time interval 2516 to 2524 ms. In order to eliminate noise, the trace corresponding to the maximum of the cross-correlation coefficients less than a threshold value was discarded. The choice of the threshold value is somewhat arbitrary and has been taken to be 0.2 in the present case. In the above example, the value at (1).274.02 is found to be maximum which is greater than the threshold value 0.2, and hence the trace#2 is retained. If the maximum value of cross-correlation coefficients had been less than 0.2, say 0.15, then we would have simply discarded trace#2 from further consideration.

Then the maximum amplitude from the window, on trace #2 was picked as the marker and the same process repeated between traces #2 and #3. This process was continued until all the traces were examined in a given section. Figure 1 shows the correlated reflection horizon using 2001 traces from a data sample. (These 2001 traces cover a distance of 50,025 meters adjoining the explosion).

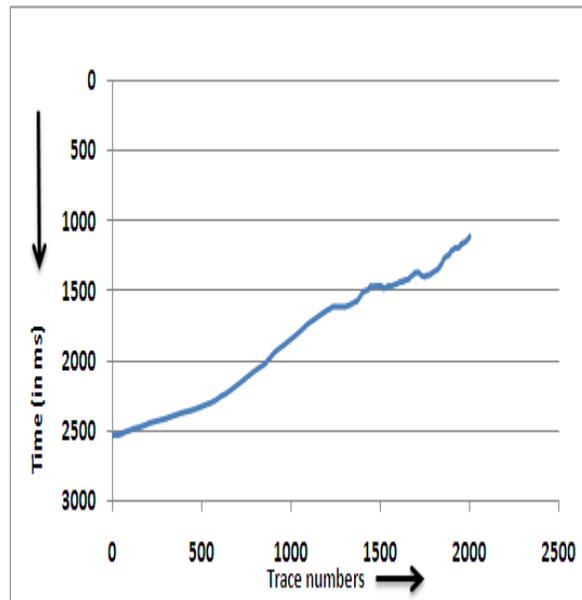


Figure 1: Correlated reflection horizon using 2001 traces.

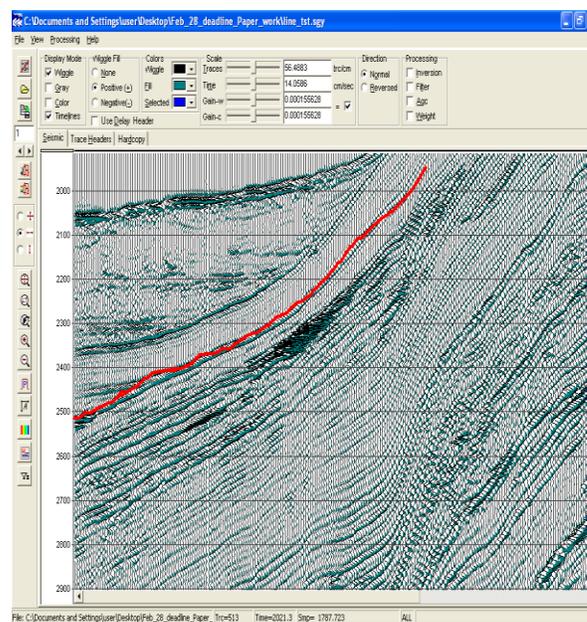


Figure 2: SEG-Y section corresponding to correlated horizon shown in Fig 1

The slope of the line in Figure 1 corresponds to the slope of the reflecting horizon A in Figure 2. The correlated horizon attains a steep upward slope. This corresponds to an interface between two types of rocks. A close similarity between the computed horizon and the one seen in SEG-Y image, establishes the correctness of the approach. In the case shown above, the maximum amplitude along with its time was chosen, after inspecting the entire trace, in order to proceed with cross-correlation. The user can also give a random time interval range, between which the trace of reflected wave is examined and maximum amplitude along

with its time is taken as its starting point to begin cross-correlation. Alternatively, the user is free to specify time on the first trace and use it to apply cross-correlation to the adjacent traces.

The cross-correlation technique was applied on several reflecting horizons from three seismic sections of anonymous geological areas. Figure 3 shows 14 horizons that have been tracked using this technique. These have been compared with the corresponding SEG-Y image shown in Figure 4. Similar results were obtained with the other two sections (not shown).

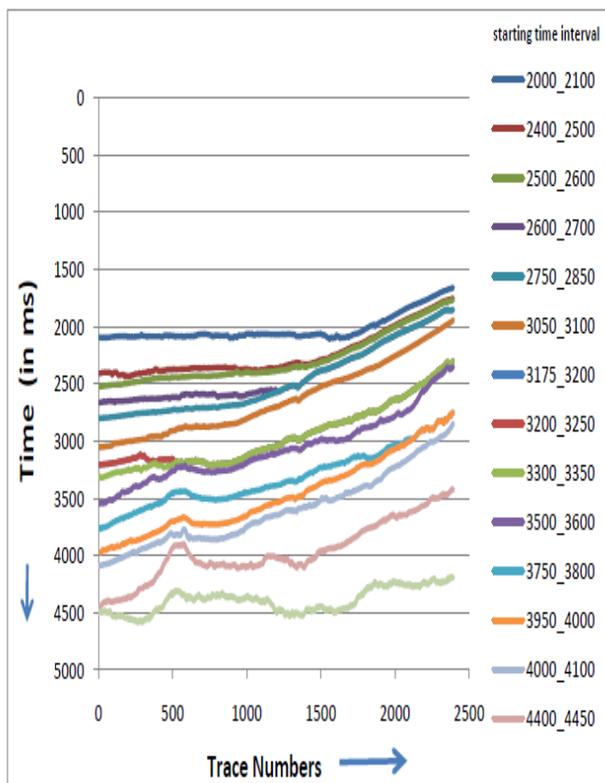


Figure 3 Reflecting horizons tracked using cross-correlation technique corresponding to horizons seen in Figure 4

Comparing the reflected horizons tracked using cross-correlation with their corresponding SEG-Y images, it was observed that technique worked well for continuous horizons. On close observation, it can be seen that while all the correlated horizons were found to be identical with those in the seismic image except the one starting between 4000 and 4100 ms which drifted away after traversing some distance. This is because of a discontinuity encountered in that horizon. This aspect has been described in greater detail in the following subsection.

4.2 Horizon tracking using secondary attributes

As mentioned earlier, MATLAB has been used to generate complexes corresponding to amplitude data. As an example consider a measured amplitude value of -0.37. Hilbert function transforms this into a complex, $-0.37-1.82j$ with $f(t) = -0.37$ and $f^*(t)=-1.82$. Use of equations (4) and (5) will yield corresponding values of instantaneous phase and reflection strength respectively.

Consider a SEG-Y section shown in Figure 4 where a fault is known to be present between trace numbers 3450 and 3650 (encircled). Let us try tracking a horizon across this interval starting at time instant 4092 ms where a strong reflection exists. We first used cross-correlation and the resultant horizon is shown in Figure 5 by curve 1. Starting with the same time instant we next tracked the instantaneous phase across the traces and the resultant horizon is also shown in the same figure by curve 2.

A comparison of the two curves with the SEG-Y section (Figure 4) clearly shows that while cross-correlation failed to track the horizon across the fault plane, the instantaneous phase identified the horizon correctly all the way. A closer look shows that the cross-correlation, as it progressed across the fault plane, picked another strong reflection horizon and moved along this horizon which was actually, a horizon above the one we started to track. But because of the rock movement downwards, after the fault it juxtaposed with the upper horizon before the 'fault'. Since, cross-correlation only correlates the strongest signal in the adjacent trace it makes no difference which horizon it picks next. However, instantaneous phase essentially remains constant for a given horizon and hence tries to identify the same horizon across the fault plane.

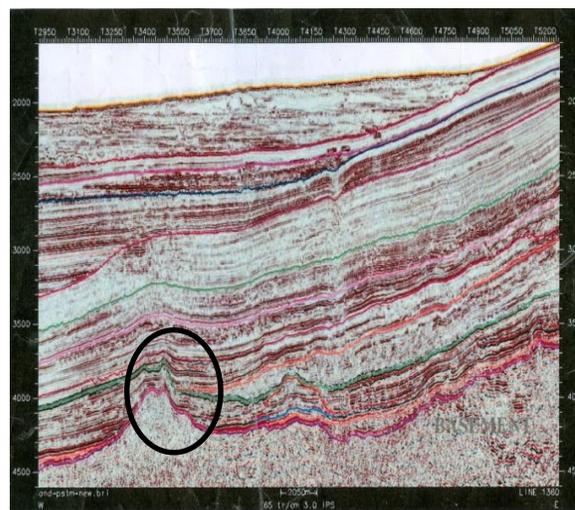


Figure 4: SEG-Y section showing a fault at bottom left

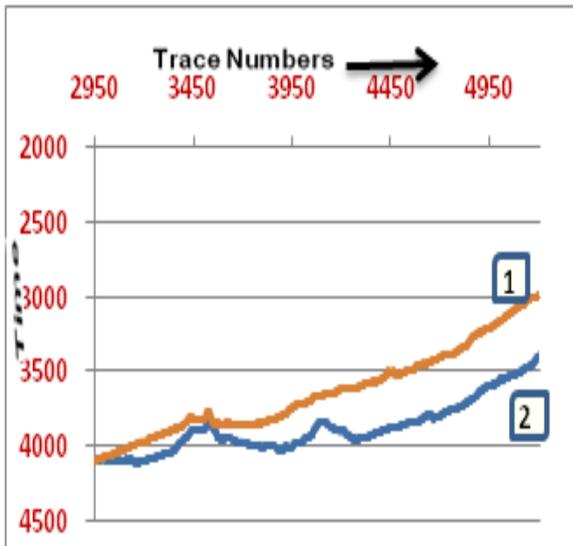


Figure 5: Tracking across a fault plane using: cross-correlation curve 1, instantaneous phase curve 2.

Reflection strength attribute was used to track the same horizon. The procedure used was similar to the one used with instantaneous phase. The starting time was same as that in case of cross-correlation (See Figure 6). As seen, in the figure, the horizon picked across the fault plane is correct one but resolution of the structure is not as good as in Figure 5.

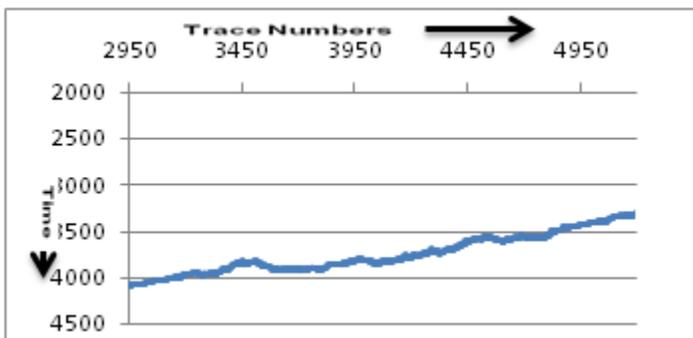


Figure 6: Tracking across a fault plane using 'reflection strength'.

4.3 Tracking fault by traversing in reverse direction (backwards)

In the close vicinity of the fault plane, there is intense heat generated because of rock shearing and this leads to changes in the rock texture and composition. Because of this change, instantaneous phase and reflection strength may also change making it difficult to delineate the precise location of the fault plane [5]. To be able to locate the fault plane more accurately, the horizon was tracked from left to right until a small distance from the fault plane on the left. The same horizon was tracked in the reverse direction, from right to

left, starting from the other end of the traces until a short distance from the fault plane on the right. The fault plane can be located more precisely in this interval.

To summarize, a total of three already interpreted seismic sections were used for validation of the present system. 14 horizons were tracked in first section and 5 and 4 respectively in the second and third sections. The system tracked all the horizons successfully. In the first section (Figure 4) there are some horizons which are monotonic, others are folded and one is faulted. All the continuous horizons were correctly tracked by cross-correlation alone and instantaneous phase and reflection strength reconfirmed their nature. However, horizons with faults, one in first section (starting at 4092 ms and shown in Figure 5) another in second section (not shown) could not be tracked correctly by cross-correlation beyond the fault planes. But when the other seismic attributes were used, complete horizons were correctly identified.

5 CONCLUSIONS

A rule-based expert system (SeisExpert) has been developed with capability to track horizons and identify faults. A graphical user interface has been provided in Visual Basic language, through which the system seeks information from the user. A lot of information can be provided from visual inspection of the seismic images which leads to a tentative interpretation of the geological structures. Analytical tools built into the 'Applications' part of the shell such as cross-correlation, instantaneous phase and reflection strength further aid in refining the interpretation. Continuous structures could be manually interpreted and reconfirmed when invoking cross-correlation program. However, discontinuities such as faults which are likely to be missed, by cross-correlation were correctly identified by the secondary attributes namely instantaneous phase and reflection strength, which are characteristic of a reflection horizon.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the management of the University of Petroleum and Energy Studies for supporting this work and granting permission to publish it.

REFERENCES

- [1] Sternbach, L.R, Unsolved mysteries of seismic interpretation: A retrospective of 20 years of TLE, The Leading Edge, Vol.21, No.10, pp 952-954, 2002.
- [2] Herron, D., Horizon auto-picking, The Leading Edge, Vol.19, No.5, pp 491-492, May,2000.

- 1.
2. [3] Admasu, F. and Toennies, K., 2006, Exploiting 3D Spatial Continuity for Robust Automatic Horizon matching across faults, Proc: 3rd International Symposium on 3D Data Processing, Visualization, and Transmission, 2006.
3. [4] Verney, P., Rainand, J.F., Perrin, M., and Thommat, M., A knowledge-based approach of seismic interpretation: horizon and dip-fault detection by means of cognitive vision, SEG, Las Vegas Annual Meeting pp 874-878, 2008.
4. [5] Taner, M.T., Koehler, F., and Sheriff, R.E., Complex seismic trace analysis, Geophysics, Vol.44. No.6, pp 1041-1063, 1979.
5. [6] Brown, A.R., Understanding seismic attributes, Geophysics Vol.66, No.1, pp 47-48, 2001.
6. [7] Chopra, S. and Marfurt K.J., Seismic attributes: a historical prospective, Geophysics Vol. 70, No.5, pp 288-328, 2005.
7. [8] Barnes, A.E., Redundant and useless seismic attributes, Geophysics, Vol.72, No.3, pp P33-P38, 2007.
8. [9] Hutchinson, Simpson, US Patent Number 5,537,320, 1997.
9. [10] Pepper, R., and Bejarano, G., Advances in Seismic fault interpretation Automation, search and discovery.net article#40169, Schlumberger Stavenger Technology, Stavenger, Norway, 2005.
10. [11] Bommel, V., and Pepper., US Patent Number 5,999,885, 1999.
11. [12] Aurnhammer, M. and Tonnie, K., Horizon correlation across faults guided by geological constraints, Proc: SPIE, Vol.4667, Electronic Imaging, 2002.
12. [13] Harrigan E, Kroh J.R, Sandham W.A, Durrani T.S, Seismic horizon picking using an artificial neural network, Proc: International conference on Acoustics, speech and signal processing 1992, March 23-26, San Fransisco, USA, Vol.3, 1992.
13. [14] Benbernou, R. and Warwick, K., A fuzzy multi-criteria decision approach for enhanced auto-tracking of seismic event, Proc: IEEE international conference on Signal processing and communications (ICSPC 2007) Nov 24-27 Dubai, UAE, 2007.
14. [15] Li, J., University of Saskatchewan, Canada, Multi-attributes pattern recognition for reservoir prediction, Canadian society of Exploration, Geophysics abstracts, 2008.
15. [16] Angeneyulu, K.S.R., Expert System introduction, article#2, Resonance, 1998.
16. [17] Duda., Hart, R.O., Konolige, P.E., and Reboh, R., A computer-based consultant for mineral exploration, Final Report of SRI project 6415, Artificial Intelligence Centre. SRI International, 1979.
- 17.
18. [18] Crain E.R., CRAIN's Petrophysical handbook, 3rd-Millennium Edition, version 3.10, Revised Dec 2009, ISBN 0-9734171-0-2 (online version), 2005.
19. [19] Wu, X., and Nyland, E., Automated Stratigraphic Interpretation of well-log data, Geophysics, Vol.52, No.12, pp 1665-1676, 1987.
20. [20] Pitas, I., Venetsanopoulos, A., AGIS: An expert system for automated geophysical interpretation of seismic images, Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'87., Vol.12, no., pp 2256 -2259, Apr 1987.
21. [21] Whitley, E.A., Expert systems: True support for the process of decision making, Proc: 1990 ACM SIGBDP Conference on Trends and directions in Expert Systems, Orlando, Florida, US, pp 123-140, 1990.
22. [22] Pazzani, M., and Iyer, R., A knowledge-based system for the management of HIV-infected patients, Proc: Intelligent Information Systems- IIS-97, pp 7-13, 1997.
23. [23] Wick, M.R., and Slagle, J.R., University of Minnesota, An explanation facility for today's expert systems, IEEE Expert: Intelligent Systems and their applications 1989, March, Vol.4, No.1, pp 26-36, 1989.
24. [24] Rattanaprateep, C., Chittayasothorn, S., A Frame-based Object Relational Expert Database System, Proc: 5th International conference on Artificial Intelligence, knowledge engineering and databases, Madrid, Spain, pp 327-332, 2007.
25. [25]Chamberlin, J., Mooney, D., Caraway, A., Talburt, J., A Rule-based expert system for check routing and cash letter management, Proc: Applied Computing Symposium 1990, April 5-6, pp 168-169, IEEE, 1990.
26. [26]Bhattacharya, S. and Raju, A.R., IIIT Hyderabad, Knowledge based Expert System for optimizing the decision making process in tender evaluation, Proc: International conference on construction and materials, IIT Kharagpur, 2003.
27. [27]Golabchi, M., A knowledge-based expert system for selection of appropriate structural systems for large spans, Asian Journal of Civil Engineering (Building and housing) Vol.9, No.2, pp 179-191, 2008.
28. [28]Tsang, C.H.K., Monde, H.C., and Woo, An object oriented Intelligent Tourist Advisor system, Intelligent Information Systems, Australian and New Zealand conference, 1996, Nov 18th -20th , Adelaide, SA, Australia, pp 6-9, 1996.
29. [29] Saatchi, MI. R., Ayienga, E.M.I., Travis, J.R., and Rippon, F., An expert system developed to assist digital electronics teaching, Engineering science and Education Journal Vol.4, No.2, pp 17-23, 1998.

30. 581,1998.
31. [30]Negnevitsky, M., A Knowledge based tutoring system for teaching fault analysis, IEEE Transactions on Power Systems, Vol. 13, No.1, pp 103-116, 1998.
32. [31] Jain, M.B., Jain, A., and Srinivas, M.B., IIIT, Hyd, A web based expert system shell for Fault diagnosis and control of power system equipment, Proc: Conference on Condition Monitoring and Diagnosis, Beijing, China, April 21-24, 2009.
33. [32] Zalis, K., Application of Expert systems in Diagnostics of High voltage insulating systems, Proc: International conference on Solid Dielectrics, Toulouse, France, July 5-9, 2004.
34. [33] Hsu, Y.Y., and Su, C.C., A Rule-based Expert system for steady-state stability analysis, IEEE, Transactions on power systems, Vol.6, No.2, pp 178-184, 1991.
35. [34] Mendis, K.S., Karunananda, A. S., Samarantunga., Ratnayake, U., An approach to the development of commonsense knowledge modeling systems for disaster management, Proc: 24th National Information Technology Conference, Colombo, pp 160-168, 2009.
36. [35] Davis, R., Austin, H., Carlbom, I., and Framley, B., The Dipmeter Advisor: Interpretation of geological signals, Proc: 7th ICAI, pp 846-849, 1981.
37. [36] Smith, R. G., and Baker, J.D., The Dipmeter Advisor System - A Case Study In Commercial Expert System Development, Schlumberger-Doll Research, Proc: 8th International Joint Conference on Artificial Intelligence, Vol.1 (IJCAI-83), 1983.
38. [37] Ketata, C., Satish M.G., Islam, M.R., Expert System Knowledge Management for Laser Drilling in the Oil and Gas Industry, Proc: CIMCA-IAWTIC'06, Vol.1, pp 930 - 938, 2005.
39. [38] Simaan, M.A., Zhang, Z., SEISIS: a rule-based system for segmentation of a seismic section based on texture, Geo-science and Remote Sensing Symposium, IGARSS, Quantitative Remote Sensing for Science and Applications, International, Vol.1, pp 461 - 463, 1995.
40. [39] His-Ho, Liu., A Rule-Based System for Automatic Seismic Discrimination, Pattern Recognition, Vol.18, No.6, pp 459-463, 1985.
41. [40] Al-garni, A.M. and Al-sari, A.M., Remote Sensing Geology and Expert Systems, Proc: American Congress on Surveying and mapping, 1994.
42. [41] Bache, T.C., Bratt, S.M., Swanger, H.J., Beall, G.W., Dashiell, F.K., Knowledge-based interpretation of seismic data using intelligent monitoring system, Bulletin of Seismological Society of America, vol:83 no:5 pp 1506 - 1526, 1993.
43. [42] Pan, J., De'Souza, G.N., and Kak, A.C., FuzzyShell: A large-scale expert system shell using fuzzy logic for uncertainty reasoning, IEEE Transactions on Fuzzy systems, Nov 98, pp 563-581,1998.
44. [43] Sheremetov, L., Batyrshin, I., Martinez, J., Rodriguez, H., Filatov, D., Fuzzy expert system for solving lost circulation problem, Applied soft computing, Vol.8, No.1, pp 14-29, 2008.
45. [44] Ali, A.G. and Haitham M., and Lababidi, S., Development of an expert system for underbalanced drilling using fuzzy logic, Journal of Petroleum Science and Engineering, Vol.31, No.1, pp 23-29, 2001.
46. [45] Kumar, B.R., Shanmugam, L, Janrathanan, S., and Santhiseela, R., Development of Expert System for the design of air-borne equipment, Digital avionics systems conference-2004, Vol.2, pp 6.c.4; 61-70, IEEE, 2004.
47. [46] McCormach, M., Neural Computing In Geophysics, Geophysics, Vol.6, No.1, pp 10-15, 1991.
48. [47] Quah, T.S., and Tan, C.L., A Neural network Expert system shell, Proc: Artificial Intelligence for Applications 10th conference Mar 1-4, San Antonia, Texas, pp 502-508, 1994.
49. [48] Spensor, C., 2004, Logic Programming Associates Ltd, UK, Flex expert system shell- documentation.

APPENDIX A: GLOSSARY OF TERMS

SEG-Y format: The SEG Y file format is one of several standards developed by the Society of Exploration Geophysicists for storing geophysical data. It is an open standard, and is controlled by the SEG Technical Standards Committee, a non-profit organization.

Reflectors/ Reflecting horizons: It is the sub-surface rock interface from where the seismic wave reflects back due to density changes of the layers.

Trace: The seismic data recorded for one channel is referred to as a seismic trace. It represents the response of the elastic wavefield to velocity and density contrasts across interfaces of layers of rock or sediments as energy travels from a source through the subsurface to a receiver or receiver array.

Edit box style: This style is useful when the user is expected to enter some text or an integer in response to the question being asked. This style presents the question along with the

text box, so that the user can enter the response to the question.

List box style: This style is used when the user is given a question and the options available to him are presented in the form of a list box and he is expected to choose any one option from the listed items

Multi-list-box style: This style of questioning is similar to the list-box style, except that the user is allowed to choose more than one items in the list box

APPENDIX B: IMPLEMENTATION NOTES

Data Storage and Retrieval

The seismic images and their corresponding amplitude data are stored in a database in MS Access and these can be accessed through the front-end using *ActiveX Data Objects (ADO)*. ADO objects help in establishing connection with database and at the same time facilitate requesting the data as record-set from database through query statements sent from the front-end. These also facilitate adding of new records by passing them from the front-end to the database for the purpose of storage and later recall.

Manual Interpretation: The seismic image, to be interpreted, is displayed on the screen, as shown in Figure B1. The questions are displayed, one at a time, on the same screen and responses are gathered from the user.

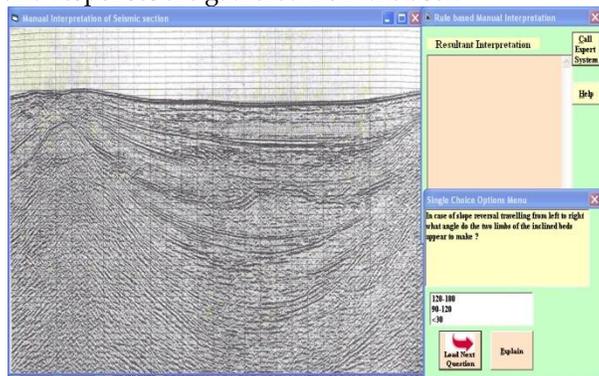


Fig B1 Seismic section (left), interpretation window (top right), questioning interface (bottom right)

The questions range from simple ones concerning the appearance of the section like: 'Where do you find the y-coordinate of the section beginning from? ('From zero', or 'At a numeric value'), to questions which are concerning a particular reflector as it moves from left to the right of the section.

The questions then begin to enquire about 'continuity of horizons', and further about 'existence of slope changes in horizons'. Further, based on the answers, other set of relevant questions are asked such as 'Is there a slope reversal present?'

Slope reversal can be indicative of presence of anti-cline/syncline/salt-dome structure. The system asks detailed questions regarding direction of slope reversal etc, to narrow down to decision regarding presence or absence of a particular structure.

Having tracked a particular horizon as one moves horizontally across the traces, the questions then concentrate on the section vertically, depth-wise, to ascertain shifts in the horizon. Questions such as 'Do you find a sudden vertical shift in the horizon' ascertain the presence of discontinuity which could be to a 'fault'. Once, a fault structure is confirmed, the questions try to ascertain the type of fault, whether normal or thrust fault, depending upon the direction of slip of the hanging block and foot wall of the proposed fault plane.

Analytical Interpretation: In order to undertake analytical interpretation, the amplitude vs time data (corresponding to the seismic image) are also called on the screen.

Cross-correlation Technique

The process starts with picking up a user-defined marker, from the first trace of the reflected wave (such as maximum amplitude value in that trace or in a specified time interval which is akin to depth interval), along with its corresponding time. Then a time window, of user-specified size (e.g. 5, 7, 9 or 11 time instants) is selected, with marker in the center of the window. This is cross correlated with the adjoining trace. These two adjacent traces within the time window represent a matrix with 2 columns and (2n-1) rows where n is the size of the window. The computations are done and the values of cross-correlation coefficients calculated. Shown below are the equations for cross-correlation coefficients for a window of five time instants.

$$a_1 = (x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 + x_5y_5) / 5 \quad (a)$$

$$a_2 = (x_1y_0 + x_2y_1 + x_3y_2 + x_4y_3 + x_5y_4) / 5 \quad (b)$$

$$a_3 = (x_1y_{-1} + x_2y_0 + x_3y_1 + x_4y_2 + x_5y_3) / 5 \quad (c)$$

$$a_4 = (x_1y_2 + x_2y_3 + x_3y_4 + x_4y_5 + x_5y_6) / 5 \quad (d)$$

$$a_5 = (x_1y_3 + x_2y_4 + x_3y_5 + x_4y_6 + x_5y_7) / 5 \quad (e)$$

In order to eliminate noise from signal, the cross-correlation coefficients lower than a certain threshold value, are discarded. The time corresponding to maximum amplitude of the adjacent trace in the matrix is taken as the centre for the next window. Each time after cross-correlation coefficient computation is done and centre for the time window is chosen, the trace from the left of matrix is removed and the adjacent trace from the right is included and the window slides ahead. This process continues till all the traces within a seismic section have been examined. Typically, a seismic

section may contain any number of traces depending upon the length of area for which the recording has been done. The time instants at the centre of the window for each seismic trace are given as output. These when plotted depict the reflecting horizon. Similar analysis is carried out to track all reflecting horizons within a given section.

Once these plots are available, questioning starts once again based on this new information. The user is asked to compare the obtained plot with the corresponding horizon from the seismic image and the responses are recorded. A close match between the obtained plot and the picked horizon indicates confirmation of observations made earlier. There is a possibility that the horizon in the image may indicate a break, which may be missing in the plot. The absence of the break in the plot may be due to the fact that, the cross-correlation technique failed to track horizons across discontinuities. In such a situation, the user is prompted through a question to use the seismic attributes, instantaneous phase and/or reflection strength to track the same horizon.

Instantaneous Phase and Reflection Strength

Starting with the amplitude vs time data, Hilbert transform generates the complex which when used with Equation (4) gives instantaneous phase as a function of time. The initiation of the process of tracking reflecting horizon using instantaneous phase takes place, with the user picking up a reasonably large value of instantaneous phase in the time window in which that particular horizon originates. Then the instantaneous phase nearest to this phase value in the adjoining trace is chosen and its corresponding time is considered as centre for the next time window. This process continues till all the traces are examined in the given section. Same procedure is used to track the reflecting horizon with the reflection strength attribute. Appropriate application programs are invited by the expert system to track the horizon and plots thus generated are displayed. The user is prompted to compare the plot with the corresponding horizon in the seismic image. A match between the two confirms the interpretation. The user is then prompted to choose the next horizon for interpretation and the same procedure is followed. Thus, the user may choose to interpret a set of horizons in a given seismic section and finally display the overall interpretation.

Help Module

The system provides context-sensitive help that can be accessed from the menu bar, help button or by pressing 'F1' key.

Expert System Shell

One sample each of a question, rule, group, rule-set and an action has been shown below. The expert system gathers information from the user through responses to the questions. Questions have been designed and added to 'Flex'

user interface, using the key word 'question', an example of which is given below.

```
question q3
On visual inspection of the section, How do
you find the reflections?
Choose one of grp_ref_type
because The dark reflections indicate density
of the strata/layer is more and if the light
reflections are found the density of the layer
is less.
```

The name of the question is **q3** and it holds the body of the question which is 'On visual inspection of the section, How do you find the reflections?'

The user is allowed to choose from a group **grp_ref_type** which provides two options 'Dark' and 'Light' that are displayed with question through a list box. The 'because' clause is used to provide question-specific instruction and explanation along with the question, to assist the user to understand and answer it appropriately.

Creating of group **grp_ref_type** with the options is shown below:

```
group grp_ref_type
'Dark', 'Light'.
```

Based on the responses to the questions appropriate rules are invoked. The response to the question is held in the name of the question, **q3** in this case.

Flex provides a very simple procedure to add rules using keyword 'rule'.

```
rule r1_q3
If ref_density is q3 and
q3='Dark'
Then
Write( 'Dark reflections indicate Density of
layers is more').
```

In the above rule, **r1_q3** is the name given to the rule. The name of the question being asked to capture data for the rule is **q3** and the statement within the **write ()** function is displayed only if the response to the given question is 'Dark'. The response to **q3** would be 'Dark' if user had chosen to answer 'Dark' from the given choices. Full-stop (.) is the command terminator.

The 'Flex' interface provides the feature to group the rules into various categories by designing specific sets using keyword 'ruleset'. Then rule sets can be appropriately chained and invoked by designing actions using 'action' keyword. For example, the ruleset **set_set1** has been designed which is made to contain three rules, **r1_q1**,

r2_q1, r3_q1. It is supposed to fire each of these rules in the first come first served manner and remove the rules fired once. The last statement indicates that when a rule misfires (invoked when the condition was not true), the process stops and exits.

```
Ruleset set_set1  
Contains r1_q1, r2_q1, r3_q1;  
Update ruleset by removing each selected rule;  
When a rule misfires do true.
```

The action **run_a1** is used to invoke the ruleset and run next action **run_a2**.

```
action run_a1;  
do restart  
and invoke ruleset set_set1 and run_a2.
```

About seventy rules have been incorporated for manual interpretation and another fifty rules for analytical interpretation.

The following code takes care of invoking Flex, from visual basic front-end. The object 'flx' of class Flex is created as follows.

Dim flx As New flex

Using this object, the functions, namely, LoadFlex(), InitGoal(), CallGoal(), ExitGoal() are called. The LoadFlex() function loads flex, and InitGoal() loads the flex file (.ksl file) holding the rules, questions, rulesets and actions. '.ksl' is the extension for the flex files. It stands for 'Knowledge Specification Language'. It is very close to natural language in its syntax and semantics. This makes it very easy to understand and code.

```
flx.InitGoal ("reconsult_rules ( '<flex code>' ) . ")
```

Then RunGoal() can be called with the main action name to be invoked.

```
flx.RunGoal ("<action_name>. ")
```

The CallGoal() function starts the execution of the expert system by putting forth the questions, firing rules, using responses to move towards the goal. ExitGoal() is a very important function to stop execution, once the rules for the current session have been fired. HaltFlex() is

used to unload the previously loaded instance of 'Flex' from memory.

Deployment of Expert System

The expert system has been packaged, using 'Package and Deployment wizard', into installable software. This wizard determines which files need to be distributed with the software, compresses them all into a cabinet file (.cab file), that acts like a setup file. This setup file can be used to install the software.

Co-Ordinate Load control and Load shedding Balance by using Microcontroller

Md. Rashidul Islam, Md.Masud Kaisar Khan, Abu Ishaque Md. Forhad

Abstract-- Energy is the basic necessity for the economic development of a country. Energy production is more costly which is impossible for us, so we should distribute the energy as maximum user's right. Now a day's load-shedding is a common buzzword in our country, for this reason the industry doesn't continue the production, the aim of our research is to continue power flow in industry and load-shed the user as a balance condition. Hence, the load shedding control system, which was earlier done manually, now-a-days, is controlled by a computer based system, developed to some more extent to direct the society to a more convenient life. This paper focuses on developing a microcontroller based procedure for controlling the load-shedding system where manual work will be minimized by selecting the feeder, substation and duration of shedding time by the user, easy to detect fault using microcontroller, To continue industrial power for effective manufacturing, over load cut for Transmission line safety.

Index Terms: Load Shedding, Microcontroller, Current Transformer.

1. INTRODUCTION

Today the world in which we live is the world of Digital Electronics. The technological advancement has taken to stage where we can do nothing without the help of sophisticated instruments like computers, phones, mobiles; wireless etc. Bangladesh is a developing country with the vigorous development of the economy in Bangladesh; Bangladesh has experienced the continuous increase of load density over the whole power distribution system. The distribution feeders may become overloaded due to load growth and substation planning and it complicates the distribution system operation in areas with high load density. There is an acute power shortage every corner of Bangladesh, it becomes unavoidable, to cut down the load from one section & supply to other section, which can be done locally or through PC remotely. This Project is a very good example of embedded system as all its operations are controlled by intelligent software inside the microcontroller, [6]. The theme of our project is to control the power grid or for the purpose of load shedding. In this project, we are using **PIC16F676** Microcontroller [2], since this controller has two ports which are more than enough for our project. Technology used as microcontrollers are the core of the today's digital circuit design in industry, this system uses it for the centralized operation and digital processing. The technology used here is embedded technology which is the future of today's modern electronics.

For developing automatic load shedding technique we have used different equipments such as microcontroller, power supply(12 volts), relay, transistor, voltage regulator, current transformer, LED, Dummy load, switch etc.

Finally we have developed a microcontroller based control system and a technique of load control for fixed load. Also, technique for overload protection system and load shedding time balancing system is shown here.

2. LOAD SHEDDING

When load increases in a system, unit governors will sense the speed change and increase the power input to the generator. Extra load will be handled by using the unused capacity of all generators operating in the system (spinning reserve)[1]. If all generators are operating at the maximum capacity (spinning reserve is 0) it is necessary to disconnect a portion of the load, equal or greater than the overload, intentionally and rapidly. As frequency is a reliable indicator of an overload situation, frequency sensitive relays can be used to disconnect a portion of the load automatically. This arrangement is referred to as Load-Shedding or Load-Saving scheme and is designed to protect system against frequency interruptions. Under frequency relays are usually installed at distribution substations where selected loads can be disconnected which will balance load and generation. The first line of these relays is set just below normal operating frequency range (59.4-59.7Hz).[5-6] When the frequency drops below this level, these relays will drop a significant percentage of system loads. If the frequency stabilizes (or increase), it

1. Md. Rashidul Islam (B.Sc in EEE), is currently in teaching at 'Internation Islamic University Chittagong'. E-mail: rashidul05eee@yahoo.com
2. Md.Masud Kaisar khan is currently studying EEE at 'Internation Islamic University Chittagong'. E-mail: masud08eee@gmail.com
3. Abu Ishaque Md.Forhad is currently studying EEE at 'Internation Islamic University Chittagong'. E-mail: forhadtaher@yahoo.com

USER © 2012

<http://www.ijser.org>

means the load drop was sufficient, but if the frequency continues dropping (with a slower rate) until it reaches the second line of relays, a second block of load is shed. This will continue until the overload is relieved or all the frequency relays have tripped.

4. MICROCONTROLLER IC

PIC16F676 - 8 BIT M WITH 2K BYTES FLASH PROGRAM

MEMORY:

Photograph:



Figure 01: Photograph of PIC16F676

3.1 Pin layout:

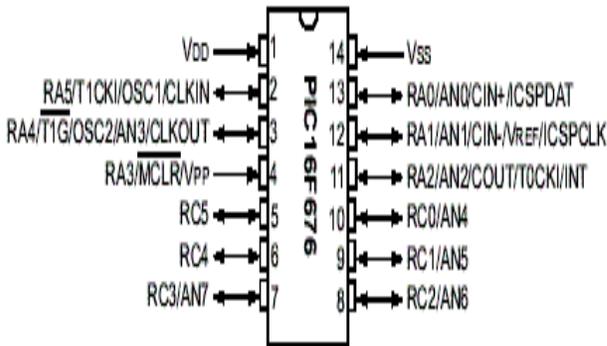


FIGURE 02: PIC16F676 LAYOUT

3.2 Features:

High-Performance RISC CPU

- Only 35 single word instructions to learn
- All instructions are 1µs (@4MHz) except for program

branches

- Operating speed: DC - 20MHz clock input

Peripheral Features

- One 8-bit timer/counter (TMR0) with 8-bit programmable presale

- 10 bit resolution

- High current source/sink for direct LED drive

- One Analog Comparators

- Capture/Compare PWM (CCP) Module

Special Microcontroller Features

- Power-On Reset

- Power-up Timer (PWRT) and Oscillator Start-Up Timer (OST)

- Brown-out Detect (BOD)

Low Power Features

- Standby Current: 1nA @ 2.0V, typical

- Operating Current: 8.5 µA @32 kHz, 2.0V, typical

- Watchdog Timer Current: 300 nA @ 2.0V, typical

- Timer1 oscillator current: 4 µA 32 kHz, 2.0V, typical

3.3 Pin description:

Pin Number	Description
1	VDD
2	RA5/T1CKI/OSC1/CLKIN
3	RA4/T1G/OSC2/CLKOUT
4	RC3/MCLR/VPP
5	RC5
6	RC4
7	RC3

8	RC2/TX/CK - Port B
9	RC1
10	RC0
11	RA2/COU/T0CKI/INT
12	RA1/CIN-/ISCPCLK
13	RA2/CIN+/ISCPDAT
14	Vss

4 VOLTAGE REGULATOR:

Voltage regulators are components that maintain a consistent voltage output. Electronic components are often made to accept only a low maximum voltage, and can be badly damaged by a power surge. Likewise, a low voltage can fail to provide enough power for the component. Voltage regulators are often responsible for maintaining a voltage within the range that the electronic component can safely accept.

A basic voltage regulator LM7805 has three legs, converts varying input voltage and produces a constant regulated output voltage. The most common part numbers start with the numbers 78 or 79 and finish with two digits indicating the output voltage. The number 78 represents positive voltage and 79 negative one. The 78XX series of voltage regulators are designed for positive input.

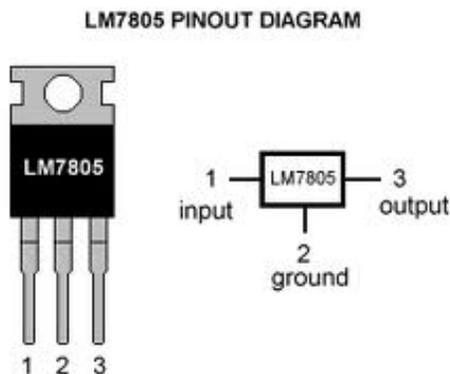


Figure 03: Standard configuration of LM7805.

5 CO-ORDINATE CURRENT CONTROL

5.1 Block diagram for co-ordinate current control:

Over load control and load balancing system is new technology in Bangladesh .Today nothing is manual to use. Automatic system is more reliable and safety for life. Here have a main generating station and one industry and two home user units. Industry is always supply by power because uninterrupted power must be use in production unit. Our generating station max power generation is 8MW.Industry is dedicated for 5 MW two users is assign for 6 MW(3+3MW) when industries is over loaded then main controller shut off home power supply. And co ordinate load shedding respect of time of total over load of industry. And industry is warning by alarm that it's over loaded.

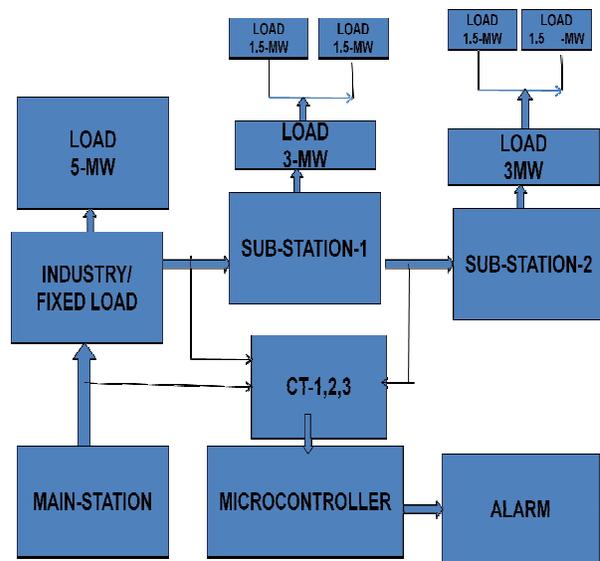


Figure 04: Block diagram for co ordinate current control

5.2 Schematic diagram for co-ordinate current control:

All transmission line is connected with CT .then all CT output is convert into DC Voltage. Which is connected with Microcontroller .Micro controller make logical decision and trigger Transistor and relay will on? if transmission line is overloaded then switch off all load. When industry signal is over to set value it will signal by alarm.

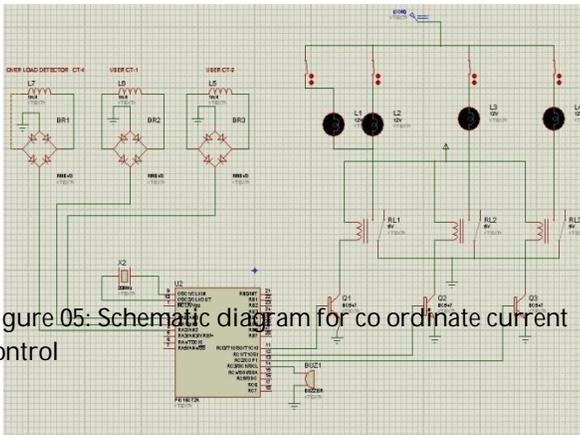
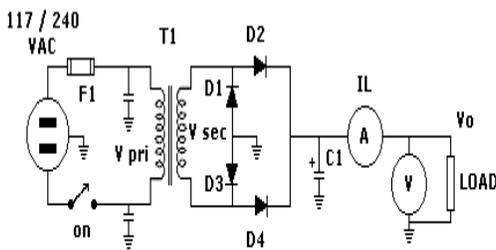


Figure 05: Schematic diagram for co ordinate current control

5.3 Power Supply:

The ac from the transformer secondary is rectified by a bridge rectifier D1 to D4 which may also be a block rectifier such as WO4 or even four individual diodes such as 1N4004 types. (See later re rectifier ratings).[3]

The principal advantage of a bridge rectifier is you do not need a centre tap on the secondary of the transformer. A further but significant advantage is that the ripple frequency at the output is twice the line frequency (i.e. 50 Hz or 60 Hz) and makes filtering somewhat easier.



$$\begin{aligned}
 V_o \text{ (no load)} &= V_{sec} \times 1.414 \\
 P_o &= V_o \times I_L \\
 \text{Load (RL)} &= V_o / I_L \\
 C1 \text{ (Vmin)} &= V_o \times 1.414 \\
 F1 \text{ (A)} &= (2 \times I) / N \text{ (N = turns ratio)} \\
 V_{sec} &= V_o / 1.414
 \end{aligned}$$

As a design example consider we wanted a small unregulated bench supply for our projects. Here we will go for a voltage of

about 12 - 13V at a maximum output current (I_L) of 500ma (0.5A). Maximum ripple will be 2.5% and load regulation is 5%.

Now the rms secondary voltage (primary is whatever is consistent with your area) for our power transformer T1 must be our desired output V_o PLUS the voltage drops across D2 and D4 ($2 \times 0.7V$) divided by 1.414.

This means that $V_{sec} = [13V + 1.4V] / 1.414$ which equals about 10.2V. Depending on the VA rating of your transformer, the secondary voltage will vary considerably in accordance with the applied load. The secondary voltage on a transformer advertised as say 20VA will be much greater if the secondary is only lightly loaded.

If we accept the 2.5% ripple as adequate for our purposes then at 13V this becomes $13 \times 0.025 = 0.325$ Vrms. The peak to peak value is 2.828 times this value. $V_{rip} = 0.325V \times 2.828 = 0.92$ V and this value is required to calculate the value of C1. Also required for this calculation is the time interval for charging pulses. If you are on a 60Hz system it is $1 / (2 \times 60) = 0.008333$ which is 8.33 milliseconds. For a 50Hz system it is 0.01 sec or 10 milliseconds.

The formula for C1 is:

$$\begin{aligned}
 C1 \text{ (uF)} &= [(I_L \times t) / V_{rip}] \times 10^6 \\
 C1 &= [(0.5A \times 0.00833) / 0.92V] \times 10^6 \\
 C1 &= 0.00453 \times 10^6 = 4529 \text{ or } 4700 \text{ uF}
 \end{aligned}$$

Remember the tolerance of the type of capacitor used here is very loose. The important thing to be aware of is the voltage rating should be at least $13V \times 1.414$ or 18.33. Here you would use at least the standard 25V or higher (absolutely not 16V).

With our rectifier diodes or bridge they should have a PIV rating of 2.828 times the V_{sec} or at least 29V. Don't search for this rating because it doesn't exist. Use the next highest standard or even higher. The current rating should be at least twice the load current maximum i.e. $2 \times 0.5A$ or 1A. A good type to use would be 1N4004, 1N4006 or 1N4008 types. These are rated 1 Amp at 400PIV, 600PIV and 1000PIV respectively. Always be on the lookout for the higher voltage ones when they are on special.

TRANSFORMER RATING - In our example above we were taking 0.5A out of the V_{sec} of 10V. The VA required is $10 \times 0.5A = 5VA$. This is a small PCB mount transformer available in Australia and probably elsewhere. This would be an

absolute minimum and if you anticipated drawing the maximum current all the time then go to a higher VA rating.

The two capacitors in the primary side are small value types and if you don't know precisely and I mean precisely what you are doing then OMIT them. Their loss won't cause you heartache or terrible problems.

6 ADVANTAGES

Here we have developed Intelligent load Control System which is more cost effective Control System. Also it is easy to detect fault using microcontroller. By this technique balance load shading time for user right can be set. To continue industrial power for effective manufacturing this technique can be more helpful. Also over load cut for Transmission line safety is possible.

7 PROSPECT OF BANGLADESH:

Bangladesh's energy infrastructure is quite small, insufficient and poorly managed. The per capita energy consumption in Bangladesh is one of the lowest (136 kWh) in the world. Noncommercial energy sources, such as wood, animal wastes, and crop residues, are estimated to account for over half of the country's energy consumption. Bangladesh has small reserves of oil and coal, but very large natural gas resources. Commercial energy consumption is mostly natural gas (around 66%), followed by oil, hydropower and coal. Bangladesh is developing country. Electricity is the major source of power for country's most of the economic activities. Bangladesh's installed electric generation capacity was 4.7 GW in 2009; only three-fourth of which is considered to be 'available'. Only 40% of the population has access to electricity with a per capita availability of 136 kWh per annum. We try to develop our country day by day. But we have no modern technology to apply our transmission line. Most of the device we buy from another country. it's very cost effective for our country. So we try to design a device to safe our transmission line and provide maximum user right to use power by controlling load shading when our production is lower from using load.

8 CONCLUSION

In this paper the importance of a timely applied load-shedding action has been reconfirmed. If the specific action is not performed in time, a more painful

load-shedding action must be performed in order to avoid a voltage collapse. Based on the discrete dynamics of LTCs, which drive voltage instability when they lose voltage regulation capability, a method for the calculation of the critical load-shedding time has been presented. Finally we can say that, by using microcontroller the control of load shedding can be made more easily. It is shown at emphasizing on shedding load of a more critical bus yields more desirable results, as expected.

9 REFERENCES:

- [1] C. Concordia, L.H. Fink, and Geroge Poullikkas, "Load shedding on an isolated system," IEEE Trans. Power Syst., vol. 10, pp. 1467-1472, Aug. 1999.
- [2] D. Guilfoyle, E. Connolly, "Distributed SCADA systems for electricity distribution control," Power Technology Int., pp. 169-172, 1998.
- [3] C.Bouchoucha, N.mansouri, M.Annabi. Adaptation of the production to the consumption of the electric power and design of the adjustment in the grid systems. GEI2002, March 2002
- [4] O. Moya, "Power system computer controlled load shedding," Electric Power System Research, pp. 165-171, April 1996.
- [5] Load Shedding – An Application Guide, John Berdy, General Electric Company Electric Utility Engineering Operation, Schenectady, N.Y.
- [6] PWang and R.Billinton, Optimum load-shedding technique to reduce the total customer interruption cost in a distribution system, IEE Proc.-Gener. Transm. Distrib., Vol. 147, No. 1, January 2000

Analysis on Adaptive Moving Objects via Robot Vision Implementations by Detection Techniques

Deepa Kumari, Shamik Tiwari, Deepika Gupta, Raina

Abstract— Building a robot is not only a passion but also a dream for most budding engineers. It is essential to make the concept of visual sensor system used in the field of robotics for identification and tracking of the objects. Identifying moving objects from a video sequence is a fundamental and critical task in many robot-vision applications. A common approach is to perform background subtraction, which identifies moving objects from the portion of a video frame that differs significantly from a background model. There are many challenges in developing a good background subtraction algorithm. First, it must be robust against changes in illumination. Second, it should avoid detecting non-stationary background objects such as moving leaves, rain, snow, and shadows cast by moving objects. Finally, its internal background model should react quickly to changes in background such as starting and stopping of vehicles.

As the name suggests, background subtraction is the process of separating out foreground objects from the background in a sequence of video frames. Background subtraction is used in many emerging video applications, such as video surveillance, traffic monitoring, and gesture recognition for human-machine interfaces. Many methods exist for background subtraction, each with different strengths and weaknesses in terms of performance and computational requirements.

Index Terms— Adaptive human-motion tracking, Background subtraction methods, Detection techniques, Frame difference, Mixture of Gaussians, Robot vision

1 INTRODUCTION

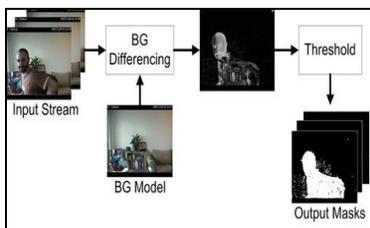
Numerous BGS algorithms and a number of post-processing techniques that aim to improve the results of these algorithms have been proposed. In this paper, we evaluated several popular, state-of-the-art BGS algorithms and examine how post-processing techniques affect their performance. The experimental results demonstrate that post-processing techniques can significantly improve the foreground segmentation masks produced by a BGS algorithm.

Let F_0 be the initial frame and F_i be the consecutive frames, where $i = 1$ to n . The pseudocode is given below:

```
If  $(F_0 - F_i) > Th$ 
{
  Then accept and process the frame
Else
  Reject the frame
}
```

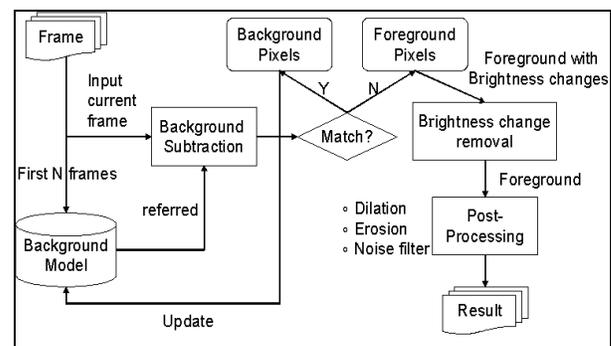
Where Th refers to Threshold value

Highlight a section that you want to designate with a certain style, then select the appropriate name on the style menu. The style will adjust your fonts and line spacing. Use italics for emphasis; do not underline.



For this evaluation, our goal is to implement three methods that were Computationally efficient enough to make the leap from MATLAB to commercial application, and good representation of background subtraction implementations in today's

video applications.



Since background subtraction is being implemented on a wide range of hardware—and thus within a wide range of computational budgets—we chose to implement methods of varying complexity;

- Low-complexity, using the frame difference method,
- Medium complexity, using the approximate median method, and
- High-complexity, using the Mixture of Gaussians method

2.1 Frame Difference

Frame difference is arguably the simplest form of background subtraction. The current frame is simply subtracted from the previous frame, and if the difference in pixel values for a given pixel is greater than a threshold T_s , the pixel is considered part of the foreground.

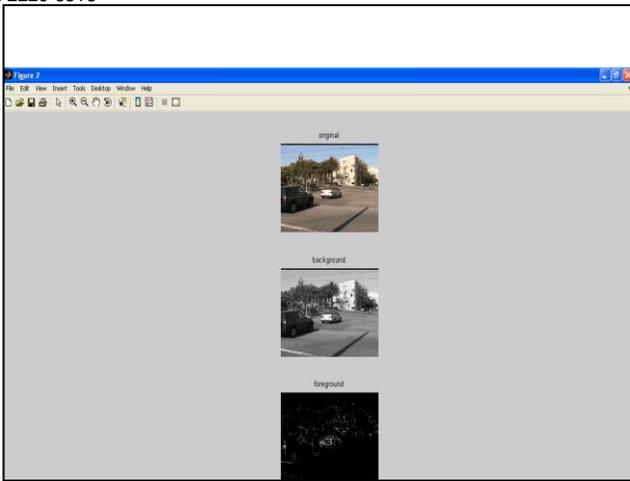


Fig 3. The frame difference method applied to the test video. Non-black pixels are foreground pixels.

1.2 Approximated Median

The approximate median method works as such: if a pixel in the current frame has a value larger than the corresponding background pixel, the background pixel is incremented by one. Likewise, if the current pixel is less than the background pixel, the background is decremented by one. In this way, the background eventually converges to an estimate where half the input pixels are greater than the background, and half are less than the background—approximately the median (convergence time will vary based on frame rate and amount movement in the scene.)

As you can see, the approximate median method does a much better job at separating the entire object from the background. This is because the more slowly adapting background incorporates a longer history of the visual scene, achieving about the same result as if we had buffered and processed N frames. Some trails behind the larger objects (the cars) can be seen. This is due to updating the background at a relatively high rate (30 fps). In a real application, the frame rate would likely be lower (say, 15 fps)

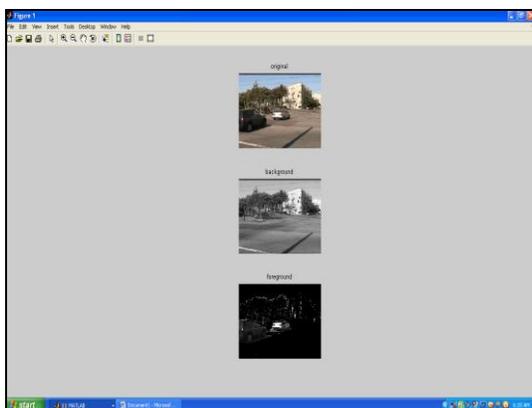
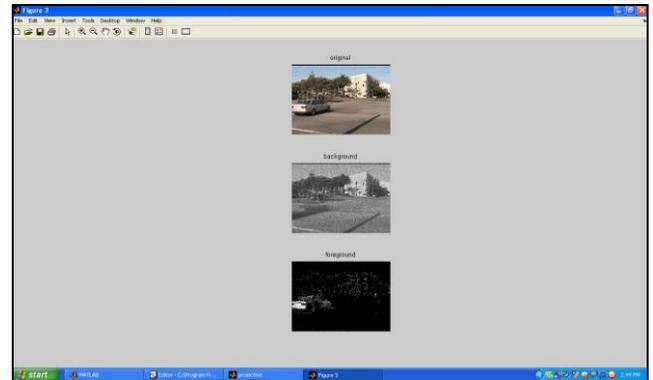


Fig 4. The approximate median method at work on the test video.

1.3 Mixture of Gaussians

Among the high-complexity methods, two methods dominate the literature; Kalman filtering and Mixture of Gaussians (MoG). Both have their advantages, but Kalman filtering gets slammed in every paper for leaving object trails that can't be eliminated. As this seems like a possible deal breaker for many applications, We went with MoG. Also, MoG is more robust, as it can handle multi-modal distributions. For instance, a leaf waving against a blue sky has two modes—leaf and sky. MoG can filter out both. Kalman filters effectively track a single Gaussian, and are therefore unimodal: they can filter out only leaf or sky, but usually not both.



In MoG, the background isn't a frame of values. Rather, the background model is parametric. Each pixel location is represented by a number (or mixture) of Gaussian functions that sum together to form a probability distributions function F.

The mean μ of each Gaussian function can be thought of as an educated guess of the pixel value in the next frame—we assume here that pixels are usually background. The weight and standard deviations of each component are measures of our confidence in that guess (higher weight & lower σ = higher confidence). There are typically 3-5 Gaussian components per pixel—the number typically depending on memory limitations.

To determine if a pixel is part of the background, we compare it to the Gaussian components tracking it. If the pixel value is within a scaling factor of a background component's standard deviation σ , it is considered part of the background. Otherwise, it's foreground

2. Implementations

Recursive techniques do not maintain a buffer for background estimation. Instead, they recursively update a single background model based on each input frame. As a result, input frames from distant past could have an effect on the current background model. Compared with non-recursive techniques, recursive techniques require less storage, but any error in the background model can linger for a much longer period of time. Most schemes include exponential weighting to discount the past, and incorporate positive decision feedback to use only background pixels for updating. Some of the representative recursive techniques are described below:

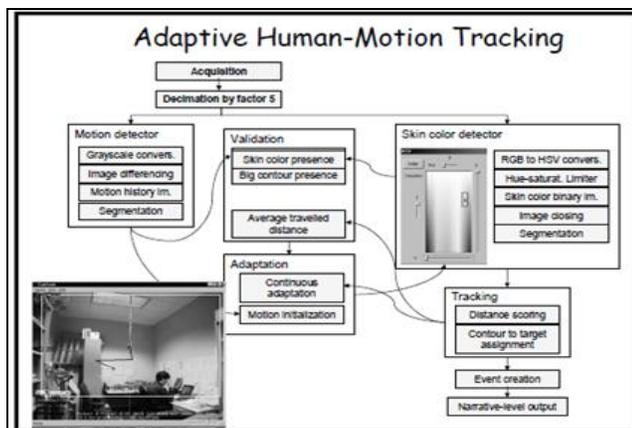


Fig 6. The processes involved in adaptive human-motion tracking.

a) Approximated median filter

Due to the success of non-recursive median filtering, McFarlane and Schofield propose a simple recursive filter to estimate the median. This technique has also been used in back-ground modeling for urban trac monitoring. In this scheme, the running estimate of the median is incremented by one if the input pixel is larger than the estimate, and decreased by one if smaller. This estimate eventually converges to a value for which half of the input pixels are larger than and half are smaller than this value, that is, the median.

b) Kalman filter

Kalman filter is a widely-used recursive technique for tracking linear dynamical systems under Gaussian noise. Many different versions have been proposed for background modeling, differing mainly in the state spaces used for tracking. The simplest version uses only the luminance intensity Karmann and von Brandt use both the intensity and its temporal derivative, while Koller, Weber, and Malik use the intensity and its spatial derivatives. The internal state of the system is described by the background intensity B_t and its temporal derivative B_t' , which are recursively updated as follows:

c) Foreground Detection

Foreground detection compares the input video frame with the background model, and identifies candidate foreground pixels from the input frame. Except for the non-parametric model and the MoG model, all the techniques introduced previously use a single image as their background models. The most commonly-used approach for foreground detection is to check whether the input pixel is significantly different from the corresponding background estimate:

d) Data Validation

We define data validation as the process of improving the candidate foreground mask based on information obtained from outside the background model. Three main limitations: first, they ignore any correlation between neighboring pixels;

second, the rate of adaption may not match the moving speed of the foreground objects; and third, non-stationary pixels from moving leaves or shadow cast by moving objects are easily mistaken as true foreground objects.

The first problem typically results in small false-positive or false-negative regions distributed randomly across the candidate mask. The most common approach is to combine morphological filtering and connected component grouping to eliminate these regions. Applying morphological filtering on foreground masks eliminates isolated foreground pixels and merges nearby disconnected foreground regions. Many applications assume that all moving objects of interest must be larger than a certain size. Connected-component grouping can then be used to identify all connected foreground regions, and eliminates those that are too small to correspond to real moving objects.

When the background model adapts at a slower rate than the foreground scene, large areas of false foreground, commonly known as "ghosts", often occur. If the background model adapts too fast, it will fail to identify the portion of a foreground object that has corrupted the background model. A simple approach to alleviate these problems is to use multiple background models running at different adaptation rates, and periodically cross-validate between different models to improve performance. Sophisticated vision techniques can also be used to validate foreground detection. Computing optical flow for candidate foreground regions can eliminate ghost objects as they have no motion. Color segmentation can be used to grow foreground regions by assuming similar color composition throughout the entire object. If multiple cameras are available to capture the same scene at different angles, disparity information between cameras can be used to estimate depth. Depth information is useful as foreground objects are closer to the camera than background. The moving-leaves problem can be addressed by using sophisticated background modeling techniques like MoG and applying morphological filtering for cleanup. On the other hand, suppressing moving shadow is much more problematic, especially for luminance-only video.

4 CONCLUSION

Although It presents a comparative study of several state-of-the-art background subtraction methods. Approaches ranging from simple background subtraction with global thresholding to more sophisticated statistical methods have been implemented and tested on different videos with ground truth. The time taken to complete an average frame of the data set is shown. The time taken varies from 0.0004 seconds to 12.7196 seconds per frame. Each of the algorithms were run 100 to calculate the average time for each frame to ensure that the operating system did not interfere or influence the speed results, apart from the Mixture of Gaussian.

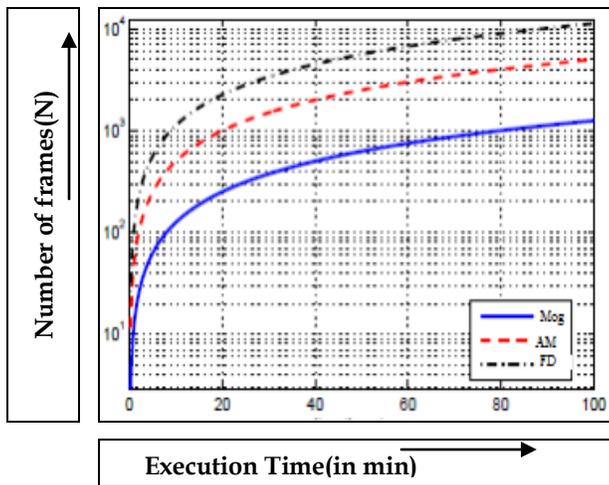


Fig 7. The graph between number of frames and execution time for the analysis of all three approached methods of background subtraction .

The only algorithm capable, within this experiment, of removing a complex background is the frame difference algorithm. This is because the model updates each frame, and checks for movement or motion through a frame, allowing for the algorithm to ignore background motion and identify foreground elements. The problem with this is that it does not handle slow moving foreground objects well. The goal is to provide a solid analytic ground to underscore the strengths and weaknesses of the most widely implemented motion detection methods. The methods are compared based on their robustness to different types of video, their memory requirements, and the computational effort they require. Most of the videos used come from state-of-the-art benchmark databases and represent different challenges such as poor SNR, multimodal background motion, and camera jitter. Overall, it helps to better understand for which type of videos each method best suits but also estimate how, sophisticated methods are better compared to basic background subtraction methods.

REFERENCES

- [1] D. Gavrila. Pedestrian detection from a moving vehicle. In ECCV '00: Proceedings of the 6th European Conference on Computer Vision-Part II, pages 37-49, London, UK, 2000. Springer-Verlag.
- [2] Wren C, Azabajejani A, Darrell T. and Pentland A. Pfindex: Real-time Tracking of the Human Body. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997 : 780-785
- [3] Haritaoglu I, Harwood D and Davis L. W4: Real-Time Surveillance of People and their Activities, IEEE Trans. Pattern Analysis and Machine Intelligence, 2000 : 809-830.
- [4] Wang Junqing, Shi Zelin, and Huang Shabai, "Detection of Moving Targets in Video Sequences". (Opto-Electronic Engineering, Dec 2005, pp. 5-8).
- [5] Ren Mingwu, and Sun Han, "A Practical Method for Moving Target Detection Under Complex Background". (Computer Engineering, Oct 2005, pp. 33-34).

- [6] Milan Sonka, Vaclav Hlavac, and Roger Boyle, "Image Processing, Analysis, and Machine Vision (Second Edition)", Posts & Telecom Press, Beijing, Sep 2003.
- [7] Zhang Yunchu, Liang Zize, Li En, and Tan Min, "A Background Reconstruction Algorithm Based on C-means Clustering for Video Surveillance", (Computer Engineering and Application, 2006, pp. 45-47).
- [8] H.G. Barrow. Parametric correspondence and chamfer matching: two new techniques for image matching. In International Joint Conference on Artificial Intelligence, pages 659-663, 1977.
- [9] D. Cunado, M. S. Nixon, and J. N. Carter, "Using gait as a biometric: via phase-weighted magnitude spectra," in Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication, 1997, pp. 95-102.
- [10] P. S. Huang, C. J. Harris, and M. S. Nixon, "Human gait recognition in canonical space using temporal templates," Proc. Inst. Elect. Eng. (IEE) Vision Image and Signal Processing, vol. 146, no. 2, pp. 93-100, 1999.
- [11] A. Johnson and A. Bobick, "A multi-view method for gait recognition using static body parameters," in Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication, 2001, pp. 301-311.

Theoretical Study of Molecular Ordering in Nematogens: 4-(4-propoxy benzilidine amino) benzoic acid

Shraddha, Navneet Chaturvedi^a, M.K. Dwivedi, S.N. Tiwari

Abstract- The peculiar changes-characteristics of mesomorphic behaviour, which occur at phase transitions, are primarily governed by intermolecular interactions acting between sides, planes and terminals of a pair of liquid crystalline molecules. In view of this fact, intermolecular interactions between a pair of 4-(4-propoxybenzilidineamino) (3BABA) molecules have been evaluated using modified second order perturbation treatment along with multicentred- multipole expansion method. Using the results of stacking, in-plane and terminal interaction energy studies, probability calculations at varied angular and positional configurations in a molecular pair of 3 BABA have been carried out employing Maxwell-Boltzmann formula. An attempt has been made to elucidate the nematogenic behaviour of 3 BABA molecules in terms of translational freedom, orientational flexibility etc.

Keywords- Intermolecular interactions, Liquid crystals, ab initio, Molecular ordering, Nematogens, Phase Transition, Interaction energy

1. Introduction

Liquid crystalline phases are stable condensed phases in which molecules pack together with order that is intermediate between the three-dimensional order of a crystalline solid and the disorder of an isotropic liquid. Liquid crystals always have partial orientational order of the molecules. Some liquid crystalline phases also have partial positional order of the molecules. The partial molecular ordering, which is the characteristic of liquid crystallinity occurs frequently in both natural and synthetic materials. Thus, liquid crystals are of considerable basic and applied interest [1-3]. In view of the key role of molecular interactions in mesogenic compounds, semi-empirical studies have been carried out by several workers.

Considerable progress has been made in understanding the liquid crystalline phases using computer simulation techniques [4-10]. Since mesogenic properties are related to molecular aggregation in a specific manner, probability distribution calculations based on interaction energy results are expected to provide information about most probable molecular aggregation, as well as tendency to retain translational and orientational order at different transition temperatures [11-15]. The present paper embodies the results of probability studies carried out in case of a nematic liquid crystal namely 4-(4-propoxy benzilidine aminobenzoic acid (3BABA) ; which passes from crystal to nematic at 463 K and nematic to isotropic phase at 545K [16].

2. Method

Net charge and corresponding dipole moment components at each of the atomic centres of the molecule have been computed by an ab-initio method, (GAMESS) with 6-31*G basis set method. Modified Rayleigh-Schrodinger second order perturbation theory along with multicentred-multipole expansion technique has been used to evaluate intermolecular interactions between a pair of 3BABA molecules. Energy minimization has been carried out, separately, for both stacking and in-plane (side-to-side and end-to-end) interactions. Accuracies up to 0.1 Å in sliding (translation) and 1° in rotation have been achieved.

The calculation of probabilities of various molecular pairs has been carried out using the Maxwell-Boltzmann formula:

$$P_i = \frac{\exp(-\beta \epsilon_i)}{\sum_j \exp(-\beta \epsilon_j)}$$

where P_i is the probability corresponding to the configuration i and $\beta = 1/kT$, where k and T are Boltzmann constant and absolute temperature respectively; and ϵ_i represents the energy of configuration i relative to the minimum energy value in a particular set for which the probability distribution is being computed. The details of mathematical formalism and optimization process can be found in literature [11-15].

3. Results and Discussion

3.1 Stacking Interactions

Table 1: Comparative probability of minimum energy stacked complexes of 3BABA with respect to translation along Z-axis corresponding to various rotational sets

Configuration	Separation (Å)	Energy kcal/mol	Probability(%)		
			300 K	463K	545K
X(0□)Y(0□)	3.5	-11.56	0.15	1.18	2.01
X(0□)Y(180□)	3.5	-15.33	84.15	71.11	65.25
X(180□)Y(0□)	3.5	-12.04	0.34	1.99	3.12
X(180□)Y(180□)	3.5	-14.31	15.10	23.36	25.34

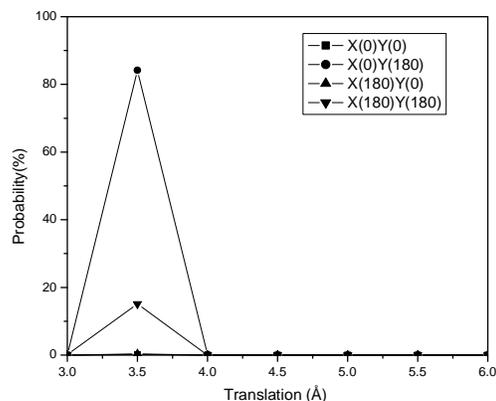


Fig 1 : Variation of probability of various molecular pairs of 3BABA with respect to translation of one of the stacked molecules along an axis perpendicular to the molecular plane and passes through the centre of mass(Z-axis) at 300K

The variation of probability with respect to translation of one of the stacked molecules along the Z-axis corresponding to four sets of axial rotations viz. X(0□)Y(0□), X(0□)Y(180□), X(180□)Y(0□) and X(180□)Y(180□) has been shown in Fig 1. It is clear from Fig 1 that the configuration X(0□)Y(180□) shows a maximum probability of 84% with energy -15.33 kcal/mole at an inter-planar separation of 3.5 Å. It is seen that for further translation, probability decreases and reduces to zero at 4.0 Å.

The configuration X(180□)Y(180□) has probability of 15.10% with energy -14.31 kcal/mole at the same inter-

planar separation 3.5 Å. The probability distribution of the rotational sets X(180□)Y(0□) and X(0□)Y(0□) shows their maxima at 3.5 Å with 0.34% and 0.15% respectively. An analysis of the relative probability of being at maxima points is shown in Table 1. It is obvious that the configuration X(0□)Y(180□) has 71.11% probability of occurrence at crystal to nematic transition temperature while configurations X(180□)Y(180□), X(180□)Y(0□) and X(0□)Y(0□) have probabilities 23.36, 1.99 and 1.18% respectively. In subsequent calculations, therefore, configuration X(0□)Y(180□) has been chosen. The effect of translation along X-axis has been studied for the configuration X(0□)Y(180□).

Table 2: Comparative probability of the minimum energy stacked complexes during translation along X-axis of 3 BABA

molecules corresponding to four selective rotations about Z-axis.

Rotation about Z-axis	Energy		Probability(%)	
	kcal/mole	300K	463K	545K
0°	-15.33	16.82	16.56	15.70
90°	-10.45	0.005	0.08	0.17
180°	-15.86	27.59	29.21	25.43
270°	-10.44	0.005	0.08	0.17

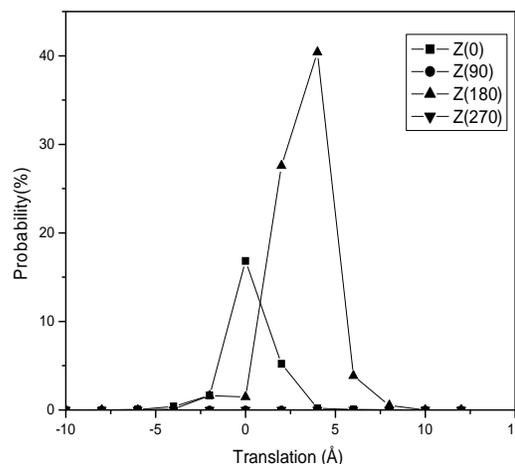


Fig 2: Variation of probability of various molecular pairs of 3BABA with respect to translation with respect to translation of one of the stacked molecules along its long molecular axis (X-axis).

The variation of probability with respect to sliding of one of the stacked molecules along the X-axis corresponding to four sets of selective rotation about Z-axis by 0, 90, 180 and 270° has been shown in Fig 2. It may be observed that the configuration with Z(180°) shows a relatively sharp peak with energy value of -15.86 kcal/mole and about 40.4% probability at the maxima. Further as shown in Table 2, the configuration Z(0°) has probability of occurrence 16.82% with energy -15.33 kcal/mole. The configuration Z(90°) and Z(270°) have nearly same probability of 0.005% respectively. Thus, the minimum energy configuration for a stacked pair of 3BABA corresponds to X(0°)Y(180°)Z(180°). This clearly indicates that perpendicularly stacked patterns of liquid crystal molecules are energetically unfavoured. Variation of probability with respect to translation of one of the stacked molecules along Y-axis is shown in

fig 3, which shows a peak at 40.5% probability value with energy -18.36 kcal/mole at the intermolecular separation of 1.2 Å. Further for translations in a narrow range, the decrement in the probability is also small (nearly 5-10%). This indicates that translation along Y-axis, in a small range, is probable at increased temperatures.

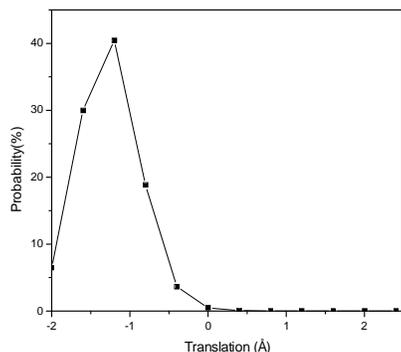


Fig 3: Variation of probability of various molecular pairs of 3 BABA with respect to translation of one of the stacked molecules along an axis perpendicular to the long molecular axis and lying in the molecular plane (Y-axis)

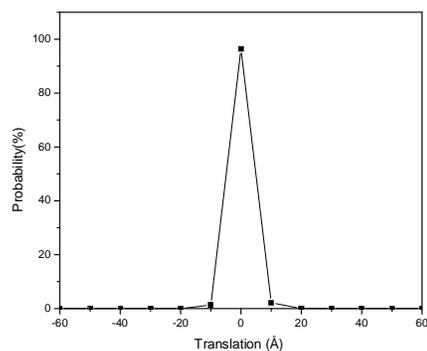


Fig 4: Variation of probability of various molecular pairs of 3BABA with respect to rotation of one of the stacked molecules along an axis perpendicular the molecular plane and passing approximately through the centre of mass of the 3BABA molecule (Z-axis).

The optimum inter- planar distance is 3.5 Å and one of the stacked possesses X(0)Y(180)Z(180) configuration.

As observed from fig 4, maximum probability 96.5% lies at perfectly aligned structure with energy value -21.19 kcal/mole at relative orientation 0°. The peak becomes sharper at decreased temperatures and, hence, makes configuration more rigid.

3.2 In-plane Interactions

The variation of probability at 300K with respect to in-plane (side-to-side) translation of one of the molecules along X-axis relative to other during in-plane interactions has been shown in Fig 5.

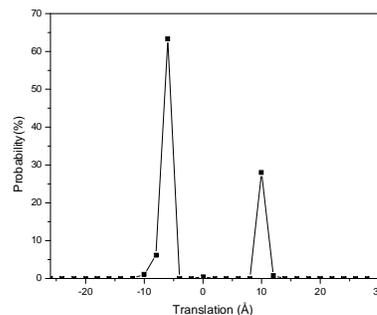


Fig5: Variation of probability of various molecular pairs of 3BABA with respect to translation of one of the molecules along the long molecular axis (X-axis) during in plane (side to side) interactions.

The intermolecular separation along Y-axis is 5.5 Å and one of the molecules possesses X(0)Y(0)Z(0) configuration. As evident from Fig 5, maximum probability (nearly 63%) occurs when one of the interacting molecules is displaced by 2.0 Å along X-axis. Similarly the transition of one of the molecules along y-axis gives rise to a probability maximum at nearly 40% value. This indicates that molecules of 3BABA a translation freedom in a small range during side-by-side interaction also.

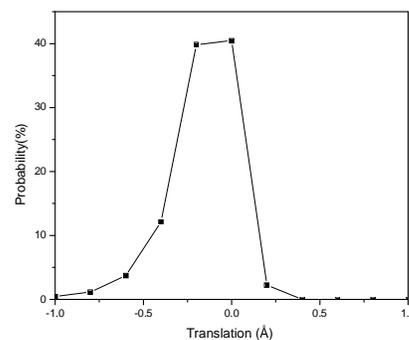


Fig:6 Variation of probability of various molecular pairs of 3BABA with respect to translation of one of the molecules along Y-axis during in plane (side to side) interactions.

3.3 Terminal Interactions

4-(4-Propoxy benzilidene amino) benzoic acid (3 BABA) is associated with -OCH₃ at one end and -COOH at the other Terminal interactions have been evaluated. All possible interacting combinations between terminal groups have been considered. It has been observed that most stable

configuration corresponds to (-OCH₃---HOOC-) interacting groups with energy -5.05 kcal/ mole and the minimum inter-group separation is 2.1 Å . Other possible interacting groups such as (-COOH---HOOC-) and (-OCH₃--CH₃O-), show energy values less than 1 kcal/mole.

Variation of probability with respect to translation along the long molecular axis in case of terminal interactions is shown in Fig 7. Maximum probability comes out to be 22% which implies a lesser possibility of the existence of end to end paired configurations of 3 BABA molecules.

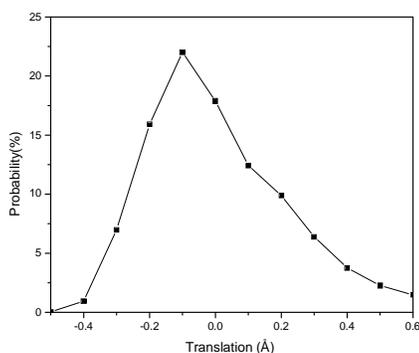


Fig 7: Variation of probability of various molecular pairs of 3BABA with respect to translation of one of the molecules along the long molecular axis (X-axis) during terminal interactions corresponding to (-OCH₃---HOOC-) interacting groups.

4. Conclusion

The present investigations reveals that 3BABA molecule has a strong preference for aligned structure at transition temperature. In a stacked molecular pair, both orientational flexibility and transitional freedom are minimal corresponding to minimum energy configuration. Other configurations show greater translational flexibility alongwith their intrinsic preference for aligned structure. Induced dipole interactions are dominant in stabilising the molecular pairs. These results favour the nematic behaviour of the 3BABA molecule.

Acknowledgements

Authors are thankful to Professor D.C.Srivastava, Head, Department of Physics DDU Gorakhpur University, Gorakhpur for providing necessary computer facility.

References

- [1] S Chandrasekhar , "Liquid Crystals", Cambridge University Press, London, 1992.
- [2] P.G.De genes, "The Physics of Liquid Crystals", Clarendon Press, Oxford, 1993.
- [3] G.W.Gray "Molecular Structure and Properties of Liquid Crystals", Academic Press, New York, 1962.
- [4] K.Tokita, K.Fuzimura, S Konodo and M Takeda Mol. Cryst. Liq. Cryst. Lett, 64, pp. 171-176, 1981
- [5] J Berges and H Perrin, "Theoretical Cxotokmatical Study of Dimethyl Azoxy Benzene" Mol. Cryst. Liq. Cryst., 113 pp.269-275, 1984
- [6] M A Bates and G R Luckhurst, "Computer Simulation Studies of Anisotropic Systems. The Density and Temperature Dependence of the Second Rank Orientational Order Parameter for the Nematic Phase of a Gay-Berne Liquid Crysta" Chem. Phys. Lett. 281, pp. 193-198, 1997.
- [7] A M Smondyrev and R A Phcovtis, "Nematic Structures in Cylindrical Cavities" Liq. Cryst. 26, pp. 235-240, 1999.
- [8] S N Tiwari, M Mishra and N K Sanyal Indian J. Phys., 76B , pp. 17-22, 2002.
- [9] S N Tiwari, "Theoretical Study of Molecular Ordering in a Nematic Liquid Crystal" Prog Crystal Growth Charact., Meter 52, USA ,pp. 150-158, 2006.
- [10] M Mishra, M K Dwivedi, M, R Shukla and S N Tiwari,), "Study of Molecular Ordering in Liquid Crystals: EBBA", Prog. Crystal Growth and Charact ., Meter 52, USA pp. 114-124, 2006.
- [11] S N Tiwari , M Mishra and R Shukla, Indian J Pure & Appl. Phys. 45, 83, 2007.
- [12] G R Luckhurst and S Romono, "Computer Simulation Study of a Nematogenic Lattice Model Based on Elastic Energy Mapping of the Pair Potential", Liq. Cryst. 26 pp.871-884, 1999.
- [13] S Sarman, "Molecular Dynamics Simulation of Thermomechanical Coupling in Cholesteric Liquid Crystal", Molecular Physics 98 pp. 27-35, 2000.
- [14] R K Bharadwaj, T J Bunning and B L Farmer Liq. Cryst. 27 (2000) 591-603.
- [15] M K Dwivedi and S N Tiwari J Mol. Liquids 158, 208 (2011)
- [16] W Maier and A Saupe (1958) Z. Naturforsch, 13A, 1958, 568.
- [17] E M Barrell and J F Jhonson Liquid Crystals and plastic Crystals, G.W. Gray and P.A. Winsor (Eds.), John Wiley & Sons, Inc., New York 254 (1974)

- Shraddha is persuing PhD from Department of Physics, D.D.U. Gorakhpur University, Gorakhpur-273009,India, Email:shraddhatri@gmail.com
- Navneet Chaturvedi is persuing PhD from Department of Electronics,Uttarakhand Technical University,Dehradun, India, Email:nav_1181@rediffmail.com
- M.K.Dwivedi has completed PhD and working in Department of Physics, D.D.U. Gorakhpur University, Gorakhpur-273009,India.
- S.N.Tiwari is working as a Proffessor in Department of Physics, D.D.U. Gorakhpur University, Gorakhpur-273009,India. Email: sntiwari123@rediffmail.com

Performance Characteristics of CI Engine Fuelled with Biodiesel and its Blends by Simulation

Sanjay Patil, Dr. M.M.Akarte

Abstract—In order to overcome the scarcity of hydrocarbon and its depletion, bio-diesel is used as a substitute for diesel or with its blend. In this paper an effort is made to develop a zero dimensional single zone thermodynamic model for compression ignition (CI) engine cycle simulation. Single Wiebe function is used to fit the experimental Mass Fraction Burned (MFB) profile to model heat released due to combustion. Heat transfer and variable specific heats are also considered. The authors of this paper have used a model to compute the performance of the CI engine powered by diesel, palm oil methyl ester (POME) and its blends with diesel. The affect of compression ratio and relative air-fuel ratio on the engine performance is analyzed. The predicted performance parameters like cumulative heat release for diesel and brake thermal efficiency with diesel, POME and various blends at 17.5:1 compression ratio are validated by comparing them with experimental results.

Index Terms— thermodynamic modeling, simulation, biodiesel, diesel engine.

1 INTRODUCTION

ENERGY is prominent requirement of present society. Internal combustion engines have been the prime movers for generating power for various applications for more than a century [1]. The increasing demand, depletion and price of the petroleum prompted extensive research worldwide on alternative energy sources for internal combustion engines. Use of straight vegetable oils in compression ignition engine for long term deteriorates the engine performance and is mainly because of higher viscosity [2-6]. The best way to use vegetable oils as fuel in compression ignition engines is to convert it into biodiesel [7]. Biodiesels such as rape seed, soybean, sunflower and Jatropha, etc. are popular substitutes for diesel [8]. In the present energy scenario efforts are being focused on use of bio diesel in compression ignition engine, but there are many issues related to performance and emission [8]. The optimum operating parameters can be determined using experimental techniques but experimental procedure will be time consuming and expensive [9]. Computer simulation [10] serves as a tool for a better understanding of the variables involved and also helps in optimizing the engine design for a particular application thereby reducing cost and time. The simulation approach allows examining the effects of various parameters and reduces the need for complex experimental analysis of the engine [11]. A validated simulation model could be a very useful tool to study engines running with new type of fuels.

The theoretical models used in the case of internal combustion engines can be classified into two main groups viz., thermodynamic models and fluid dynamic models. Thermodynamic models are mainly based on the first law of thermody-

namics and are used to analyze the performance characteristics of engines. The pressure, temperature and other required properties are evaluated with respect to crank angle or with respect to time. On the other hand fluid dynamics models are also applied for the simulation of combustion process. These governing equations in these models are partial differential equations which deal with spatial distribution of the gas velocity, temperature and composition within the engine cylinder. The solutions for those equations are tedious. Large computer memory and more computational time is needed to solve the partial differential equations. A zero-dimensional single-zone model as compared with multi-zone models is much simpler, quicker and easier to run. [12, 13] and it is capable of predicting engine performance and fuel economy accurately with a high computational efficiency [14]. Hence a zero-dimensional single-zone model is developed based on first law of thermodynamics and programmed in MATLAB to predict the engine performance for varying blend, compression ratio and relative air-fuel ratio.

2 DESCRIPTION OF MATHEMATICAL MODELING

2.1 List of symbols

r = compression ratio.

L = length of connecting rod (mm).

B = bore diameter (mm).

V_{disp} = displacement volume (m^3).

θ = angular displacement in degrees with respect to bottom dead center (BDC).

θ_s = crank angle at the start of combustion.

γ = specific heat ratio.

P = pressure (bar).

- Sanjay patil, Guru Nanak Dev Engineering College, Bidar- India
sanjaypatil1021@yahoo.co.in
- MMAkarte, National Institute of Industrial Engineering Mumbai-India. mmakarte@rediffmail.com

V = volume (m³).

m_c = number of moles of carbon in one mole of fuel.

m_h = number of moles of hydrogen in one mole of fuel.

m_o = number of moles of oxygen in one mole of fuel.

m = mass of the charge (kg).

h_c = coefficient of heat transfer due to convection (W/m².K).

A = interior surface area of cylinder (m²).

T = instantaneous gas temperature (Kelvin).

T_w = cylinder wall temperature (Kelvin).

R = universal gas constant (kJ/kmole.kelvin).

C_m = piston mean speed (m/s).

U = internal energy.

H = enthalpy.

C_p = specific heat at constant pressure (kJ/kg.kelvin).

C_v = specific heat at constant volume (kJ/kg.kelvin).

n = wiebe's shape factor.

$\Delta\theta$ = combustion duration in crank angle (degrees).

Q_r = heat released per cycle (kJ).

$\frac{dQ_r}{d\theta}$ = rate of heat released during combustion (kJ/°CA).

$\frac{dQ_h}{d\theta}$ = rate of heat transfer (kJ/°CA).

$\frac{dw}{d\theta}$ = rate of work done.

$\frac{du}{d\theta}$ = rate of change of internal energy.

$\frac{dV}{d\theta}$ = incremental change in cylinder volume (m³/°CA).

$\frac{dT}{d\theta}$ = rate of temperature change (Kelvin /°CA).

2.2 Energy balance equation

According to the first law of thermodynamics, the energy balance equation for the closed cycle is

$$m \frac{du}{d\theta} = \frac{dQ_r}{d\theta} - \frac{dw}{d\theta} \quad (1)$$

The heat term (rate of heat release) can be split into the heat released due to combustion of the fuel and the heat transfer that occurs to the cylinder walls or from the cylinder walls to gases. The equation (1) can be written as

$$m \frac{du}{d\theta} = \frac{dQ_r}{d\theta} - \frac{dQ_h}{d\theta} - \frac{dw}{d\theta} \quad (2)$$

Replacing the work transfer by $p \frac{dV}{d\theta}$ or by the ideal gas law

$$PV = mRT \frac{dV}{d\theta}, \text{ rate of heat transfer by } h_c = A(T - T_w)$$

and the internal energy can be related to specific heat through the relationship

$$\frac{du}{d\theta} = C_v \frac{dT}{d\theta}$$

Upon simplification we get equation (2) as

$$\frac{dT}{d\theta} = \frac{1}{mC_v} \frac{dQ_r}{d\theta} - \frac{h_c A(T - T_w)}{mC_v} - \frac{RT}{C_v V} \frac{dV}{d\theta} \quad (3)$$

Solving above equation by Range-kutta fourth order algorithm, the temperature at various crank angles during combustion can be calculated.

2.3 Cylinder volume at any crank angle

The slider crank angle formula is used to find the cylinder volume at any crank angle [10]

$$V(\theta) = V_{disp} \left[\frac{r}{r-1} - \frac{1-\cos\theta}{2} + \frac{1}{2} \sqrt{\left(2\frac{L}{S}\right)^2 - \sin^2\theta} \right] \quad (4)$$

2.4 Compression and Expansion strokes

The compression stroke starts from the moment the inlet valve closes (IVC) to the moment the fuel injection starts. The expansion stroke starts from the moment combustion ends to the moment the exhaust valve opens (EVO). During these processes the temperature and pressure at each step are calculated using ideal gas equation and an isentropic process [15].

$$P_{i+1} = P_i \left(\frac{V_i}{V_{i+1}} \right)^\gamma \quad (5)$$

$$T_{i+1} = T_i \left(\frac{V_i}{V_{i+1}} \right)^{\gamma-1} \quad (6)$$

$$\gamma = \frac{C_p}{C_v} \quad (7)$$

Equations 5 and 6 are used to calculate the pressure and temperature at each cylinder volume. The C_p and C_v are specific heat at constant pressure and specific heat at constant volume of the charge are calculated based on charge composition and temperature using polynomial equation.

$$C_p = \frac{\sum N_i C_{p_i}}{\sum N_i} \quad (8)$$

Where N_i is the number of moles of any gases CO (carbon monoxide), CO₂ (carbon dioxide), H₂O (water), N₂ (nitrogen) and O₂ (oxygen).

$$C_v = C_p - R \quad (9)$$

2.5 Combustion Process

Heat release due to combustion is calculated by using single Weibe's heat release correlation [16, 17].

$$\frac{dQ_r}{d\theta} = a(n+1) \left(\frac{Q_r}{\Delta\theta} \right)^n \left(\frac{\theta - \theta_s}{\Delta\theta} \right)^n \exp \left[-a \left(\frac{\theta - \theta_s}{\Delta\theta} \right)^{n+1} \right] \quad (10)$$

Where a and n are wiebe constants. These constants are different at every operating condition and they have to be such that the simulated Mass Fraction Burned profile (MFB) matches closely with experimental MFB profile. These constants are obtained by fitting the Weibe's heat release correlation to the experimental MFB profile by least square method. Prior knowledge of actual overall equivalence ratio is necessary because the fuel/air equivalence ratio depends on the amount of fuel injected inside the cylinder, from which the mass of fuel admitted can be calculated [18].

2.6 Heat transfer

The convective heat transfer between gases and cylinder wall is considerable and hence it directly affects the engine performance. The convection heat transfer in kJ/degree crank angle is given by

$$\frac{dQ_h}{d\theta} = h_c A (T - T_w) \quad (11)$$

Where heat transfer coefficient due to convection (h_c) is given by Hohenberg equation [19].

$$h_c = \frac{130 P^{0.8} (C_m + 1.48)^{0.8}}{V^{0.06} T^{0.4}} \quad (12)$$

2.7 Ignition delay

An empirical formula, developed by Hardenberg and Hase [20] is used for predicting Ignition delay in crank angle degrees.

$$ID = (0.36 + 0.22C_m) \exp \left[E_A \left(\frac{1}{RT} - \frac{1}{17,190} \right) \left(\frac{21.2}{P - 12.4} \right)^{0.63} \right] \quad (13)$$

Where ID = ignition delay period.

E_A is apparent activation energy

2.8 Gas properties calculation

A hydrocarbon fuel can be represented by $C_x H_y O_z$. The required amount of oxygen Y_{cc} for combustion per mole of fuel is given by:

$$Y_{cc} = m_c + 0.25m_h - 0.5m_o \quad (14)$$

The minimum amount of oxygen required (Y_{min}) for combustion per mole of fuel is

$$Y_{min} = Y_{cc} - 0.5m_c$$

The gaseous mixture properties like internal energy (U), enthalpy (H) specific heats at constant pressure (C_p) and constant volume (C_v) depend on the chemical composition of the reactant mixture, pressure, temperature and combustion process and can be calculated using following equations.

$$U(T) = A + (B - R) * T + C * \ln(T) \quad (15)$$

$$H(T) = A + B * T + C * \ln(T) \quad (16)$$

$$C_p(T) = B + \frac{C}{T} \quad (17)$$

$$C_v(T) = (B - R) + \frac{C}{T} \quad (18)$$

Here A, B and C are the coefficients of the polynomial equation.

2.9 Friction losses

Total friction loss calculated by the equation [21].

$$FP = C + 1.44 \frac{C_m * 1000}{B} + 0.4(C_m)^2 \quad (19)$$

Where FP is total friction power loss and C is a constant, which depends on the engine type,

$C = 75$ kPa for direct injection engine.

3 METHODOLOGY

3.1 Simulation

A thermodynamic model based on the First law of thermodynamics has been developed. Simulation program is made to model the compression, combustion and expansion processes as closed cycle period and the exhaust and intake processes by the gas exchange processes using the control volume approach. Single Wiebe function is used to fit the experimental mass fraction burned profile.

The molecular formula of diesel fuel is taken as C₁₀ H₂₂ and biodiesel is approximated as C₁₉H₃₄O₂. A computer program in matlab has been developed for the numerical solution of the

equations used in the thermodynamic model described in Section 2. This computes pressure, temperature, mass fraction burned and the performance parameters like brake thermal efficiency, peak pressure and MFB profile for the fuels considered for analysis. Fuels considered for analysis are diesel, blend of 40% biodiesel & 60% diesel and biodiesel (B0, B40 and B100).

3.2 Experimental

A stationary single cylinder, 4 stroke, water cooled diesel engine developing 5.2 KW at 1500 rpm is used for investigation. The technical specifications of the engine are given in Table 1. The fuel properties are determined using standard procedure and tabulated in table 2. The author [22] has used the experimental setup to collect data over the various blends

TABLE1. SPECIFICATIONS OF ENGINE

Sl No	Parameter	Specification
1	Type	Four stroke DI single cylinder diesel
2	Software used	Engine soft
3	Nozzle opening	200 bar
4	Rated power	5.2 KW @1500 rpm
5	Cylinder diameter	87.5 mm
6	Stroke	110 mm
7	Compression ratio	17.5:1
8	Injection timing	23 degree before

TABLE.2. PROPERTIES OF DIESEL AND POME

Sl No.	Properties	Diesel	POME(B100)
1	Viscosity in cst at(30°C)	4.25	4.7
2	Flash point(°C)	79	190
3	Fire point(°C)	85	210
4	Carbon residue (%)	0.1	0.64
5	Calorific value(kj/ kg)	42000	36000
6	Specific gravity(at25°C)	0.830	0.880

as detailed in table 3: below.

TABLE 3. DETAILS OF BLEND

B0	100% diesel
B20	Blend of 20% POME and 80% diesel
B40	Blend of 40% POME and 60% diesel
B60	Blend of 60% POME and 40% diesel
B80	Blend of 80% POME and 20% diesel
B100	100% POME

The cylinder pressure data is recorded by using piezoelectric transducer for 80 cycles. The average of data for 80 cycles is computed to evaluate mass fraction burned profile and combustion duration within the framework of first law of

thermodynamics.

4 RESULTS AND DISCUSSION

4.1 Effect of compression ratio on

4.1.1 Peak pressure

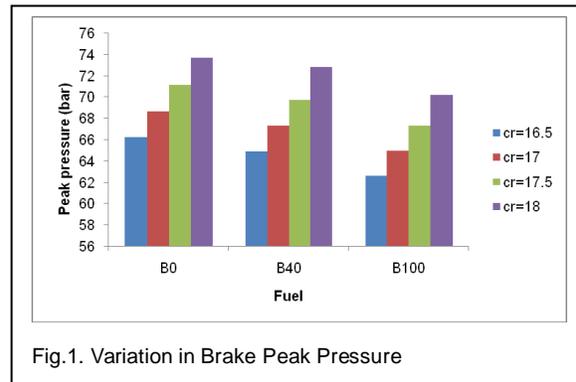


Fig.1. Variation in Brake Peak Pressure

Figure.1 shows variation in peak pressure with respect to compression ratio for different test fuels. With increase in compression ratio the air-fuel mixture compressed to high pressure and temperature which results in better combustion and hence high peak pressure. The above figure shows the same trend for all test fuels. The peak pressure is reduced as the percentage of biodiesel in the blend is increased; this is due to diffusion phase of combustion is predominant as compared to diesel and lower calorific value of biodiesel.

4.1.2 Brake thermal efficiency

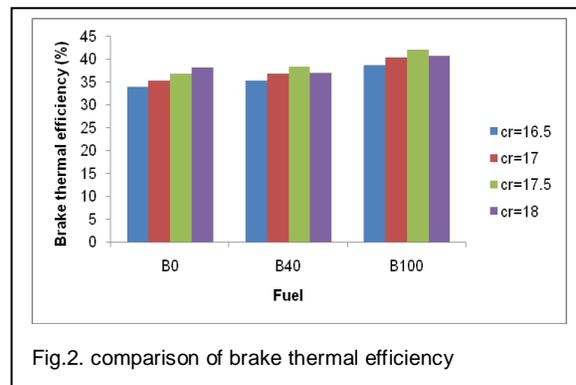


Fig.2. comparison of brake thermal efficiency

Fig. 2 shows the comparison of brake thermal efficiency with respect to compression ratios for different test fuels. With increasing compression ratio the brake thermal efficiency is increased for all the test fuels. It is observed that the brake thermal efficiency is higher with B100 than that of diesel fueled engine due to the presence of oxygen molecule in the biodiesel enhances combustion.

4.2 Effect of Relative air-fuel ratio on

4.2.1 Peak Pressure

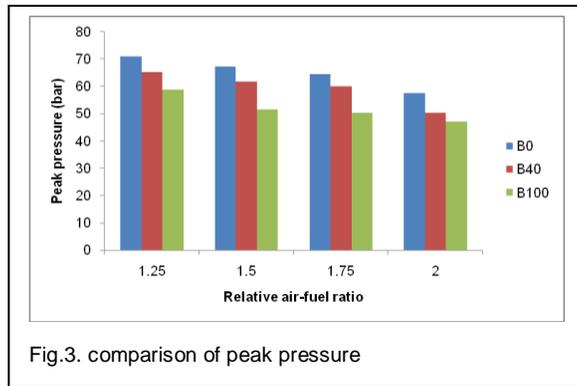


Fig.3. comparison of peak pressure

Fig. 3 shows the comparison of peak pressure with respect to relative air -fuel ratio for different test fuels. Lean mixture results in less heat release and less peak pressure whereas rich mixture results in higher peak pressure. The peak pressure is decreasing as relative air-fuel ratio is increasing. Lean mixture dilutes the charge and hence lower temperature and pressure are achieved at higher relative air-fuel ratios.

4.2.2 Brake thermal efficiency

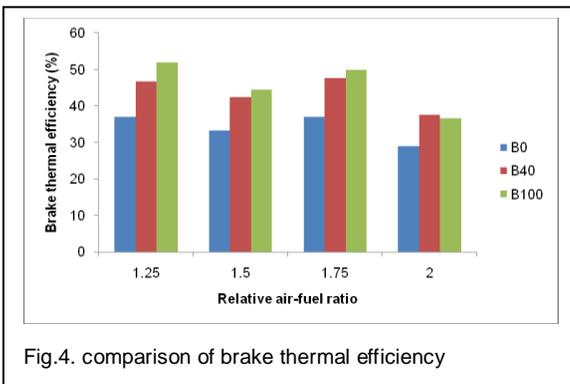


Fig.4. comparison of brake thermal efficiency

Fig. 4 shows the comparison of brake thermal efficiency with respect to relative air -fuel ratio for different test fuels. It can be observed from the fig.6 that the increase in relative air-fuel ratio dilutes the concentration of the charge and in complete combustion hence reduces the brake thermal efficiency. The same trend is observed for all the test fuels.

5 MODEL VALIDATION

With the help of developed model theoretical results are predicted for brake thermal efficiency for all test fuels and cumulative heat released for diesel at full load. The same are compared with that of experimental results on the basis of [22] for validation. The figures below highlight the features.

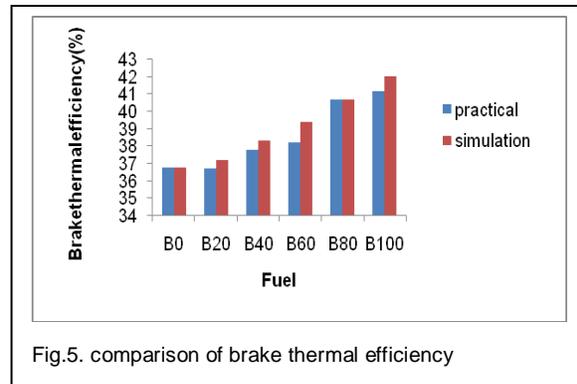


Fig.5. comparison of brake thermal efficiency

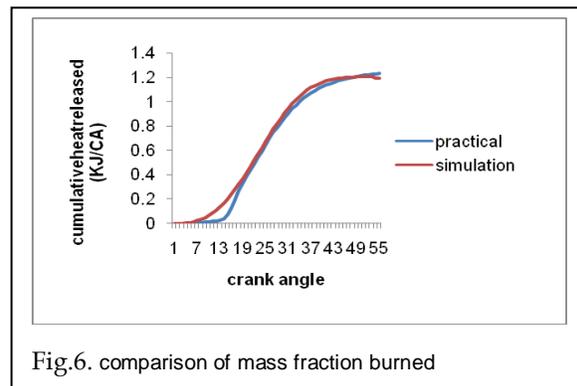


Fig.6. comparison of mass fraction burned

Predicted full load brake thermal efficiency when engine is fuelled with B0, B20, B40, B60, B80 and B100 and mass fraction burned profile for diesel are compared with experimental results are found in closer approximation and hence the developed simulation model has been proved to be reliable and adequate for the proposed objectives.

6 CONCLUSIONS

A thermodynamic model is developed for analyzing the performance characteristics of the compression ignition engine. The model is developed in such way that it can be used for characterizing any hydrocarbon-fueled engines, viz., diesel or biodiesel or their blends. The modeling results showed that, with increase in compression ratio peak pressure and brake thermal efficiency are increased. With increase in relative air-fuel ratio, the above parameters are decreased. The performance characteristics of the engine follow the same trend for all test fuels. The predicted results are compared with the experimental results of the engine fueled by B0, B40 and B100. This model predicted the engine performance characteristics in closer approximation to that of experimental results. Hence, it is concluded that this model can be used for the prediction of the performance characteristics of the compression ignition engine fueled by any type of hydrocarbon fuel.

REFERENCES

[1] Jamil Ghojel, Damon Honnery. Heat release model for the combustion of diesel oil emulsions in DI diesel engines. Applied Thermal Engineering 25 (2005)

2072-2085.

- [2] O.M. I. Nwafor & G. Rice. Performance of Rapeseed Oil Blends in a Diesel Engine. *Applied Energy*. Vol. 54, No. 4, pp. 345-354, 1996.
- [3] Vellguth G. Performance of vegetable oil and their monoesters as fuels for diesel engines. SAE 831358, 1983.
- [4] Tadashi, Young. Low carbon build up, low smoke and efficient diesel operation with vegetable oil by conversion to monoesters and blending of diesel or alcohols. SAE 841161, 1984.
- [5] Recep Altin, Selim C etinkaya, Huseyin Serdar, Yucesu, The potential of using vegetable oil fuels as fuel for diesel engines *Energy Conversion and Management* 42 (2001) 529-538 diesel engine, *Applied Energy* 86 (2009) 106-112.
- [6] Murugesan, C.Umarani, R.Subramanian, N.Nedunchezian. Bio-diesel as an alternative fuel for diesel engine- A review. *Renewable and Sustainable Energy Reviews* 13(2009)653-662.
- [7] N.R. Banapurmath, P.G. Tewari, R.S.Hosmath. Performance and emission characteristics of a DI compression ignition engine operated on Honge, Jatropa and sesame oil methyl esters. *Renewable Energy* 33 (2008) 1982-1988.
- [8] T. Ganapathy, K. Murugesan, R.P. Gakkhar, "Performance optimization of Jatropa biodiesel engine model using Taguchi approach" *Applied Energy* (2009).
- [9] T. Ganapathy, K. Murugesan, R.P. Gakkhar "Performance optimization of Jatropa biodiesel engine model using Taguchi approach"
- [10] Ganesan, V., Computer simulation of Compression-Ignition engine processes, University Press(India) Ltd., Hyderabad, India, 2000.
- [11] Udarapandian, "Performance and Emission Analysis of Bio Diesel Operated CI Engine" *Journal of Engineering, Computing and Architecture* Volume 1, Issue 2, 2007.
- [12] G.H. Abd Alla, A.A. Soliman, O.A. Badar, M.F. Adb Rabbo, Combustion quasi-two zone predictive model for dual fuel engines, *Energy Conversion and Management* 42 (2001) 1477-1498.
- [13] P.A. Lakshminarayanan, Y.V. Aghav, A.D. Dani, P.S. Mehta, Accurate prediction of the heat release in a modern direct injection diesel engine, *Proceedings of the Institute of Mechanical Engineers* 216 (2002) 663-675.
- [14] (Krieger and Borman, 1966; Foster, 1985; Assanis and Heywood, 1986)
- [15] Jamil Ghojel, Damon Honnery. Heat release model for the combustion of diesel oil emulsions in DI diesel engines. *Applied Thermal Engineering* 25 (2005) 2072-2085.
- [16] V. Caika, J. Krammer, R. Tatshi, B. Weissbacher, An integral 1D/3D workflow for analysis and optimization of injection parameters of a diesel engine, SAE technical paper, 2001, (2001-28-0004).
- [17] Lydia CDS. Simulation of the thermodynamic processes in diesel cycle internal combustion engines. SAE 931899, 1993.
- [18] P. Arque's, *La combustion: Inflammation, combustion, pollution, applications*, Ellipses, Paris, 2004, ISBN 2-7298-2037-X, p. 304.
- [19] Hohenberg GF. Advanced approaches for heat transfer calculations. SAE 790825, 1979.
- [20] J.B. Heywood, *Internal Combustion Engines Fundamentals*, Mc Graw Hill, 1988, ISBN 0-07-100499-8.
- [21] Shroff, H. D., Hodgetts, D., *Simulation and Optimization of Thermodynamic Processes of Diesel Engine*, SAE 740194, 1974.
- [22] Sanjay Patil, Dr. M.M. Akarte and Nandkishore D. Rao., Experimental Investigation on CI Engine Performance Fueled with Biodiesel and its Blends, *International Journal of Engineering Research and Technology* Volume 4, Number 3 (2011), pp. 319-327.

Setting CPU Affinity in Windows Based SMP Systems Using Java

Bala Dhandayuthapani Veerasamy, Dr. G.M. Nasira

Abstract—Multithreaded program involves multiple threads of control within a single program on single or multiple environments. In multithreaded programming model, a single process can have multiple, concurrent execution paths on CPUs. *Thread affinity* benefit a thread to run on a specific subset of processors. Multithreaded programming is written in many programming languages with an improvement of setting an affinity to threads. Java supports to develop multithreaded programming, while it does not contain any method to set an affinity for threads on CPU. This paper articulates the method to setting CPU affinity for threads in windows based SMP systems using Java.

Index Terms— Affinity mask, JNA, Multithread, Thread, Windows programming

1 INTRODUCTION

MOST concurrent applications [3] are organized around the execution of tasks: abstract, discrete units of work.

Dividing the work of an application into tasks simplifies program organization, facilitates error recovery by providing natural transaction boundaries, and promotes concurrency by providing a natural structure for parallelizing work. The first step in organizing a program around task execution is identifying sensible task boundaries. Ideally, tasks are independent activities: work that doesn't depend on the state, result, or side effects of other tasks. Independence facilitates concurrency, as independent tasks can be executed in parallel if there are adequate processing resources. For greater flexibility in scheduling and load balancing tasks, each task should also represent a small fraction of your application's processing capacity.

A multithreaded program [5] contains two or more parts that can run concurrently. Each part of such a program is called a thread, and each thread defines a separate path of execution. Each thread in a multithreaded process [9] can be dispatched to a different processor in a multiprocessor system. This collaboration across multiple processors improves single-application performance. Multithreading enables [5] you to write very efficient programs that make maximum use of the CPU, because idle time can be kept to a minimum. Threads exist in several states. A thread can be running. It can be ready to run as soon as it gets CPU time. A running thread can be suspended, which temporarily suspends its activity. A suspended thread can then be resumed, allowing it to pick up where it left off. A thread can be blocked when waiting for a resource. At any time, a thread can be terminated, which halts

its execution immediately. Once terminated, a thread cannot be resumed. Thread priority determines how that thread should be treated with respect to the others. Thread priorities are integers that specify the relative priority of one thread to another. As an absolute value, a priority is meaningless; a higher-priority thread doesn't run any faster than a lower-priority thread if it is the only thread running. Instead, a thread's priority is used to decide when to switch from one running thread to the next. This is called a context switch. Problems can arise from the differences in the way that operating systems context-switch thread of equal priority. Because multithreading introduces an asynchronous behavior to your programs, there must be a way for you to enforce synchronicity when you need it. The monitor is a control mechanism first defined by C.A.R. Hoare. You can think of a monitor as a very small box that can hold only one thread. Once a thread enters a monitor, all other threads must wait until that thread exits the monitor.

Windows supports concurrency among processes because threads in different processes may execute concurrently. Moreover, multiple threads within the same process may be allocated to separate processors and execute simultaneously. A multithreaded process achieves concurrency without the overhead of using multiple processes. Threads within the same process can exchange information through their common address space and have access to the shared resources of the process. Threads in different processes can exchange information through shared memory that has been set up between the two processes. Multiprocessor programming is challenging because modern computer systems are inherently asynchronous: activities can be halted or delayed without warning by interrupts, preemption, cache misses, failures, and other events. These delays are inherently unpredictable, and can vary enormously in scale: a cache miss might delay a processor for fewer than ten instructions, a page fault for a few million instructions, and operating system preemption for hundreds of millions of instructions.

The general-purpose process and thread facility must support the particular process and thread structures of the

- Bala Dhandayuthapani Veerasamy is currently a Research Scholar in Information Technology, Manonmaniam Sundaranar University, India. E-mail: dhanssoft@gmail.com
- Dr. G.M. Nasira is currently working as an Assistant Professor-Computer Applications, Department of Computer Science, Govt Arts College (Autonomous), Salem -636007, Tamilnadu, India. E-mail: nasiragm99@yahoo.com

various OS clients. It is the responsibility of each OS subsystem to exploit the Windows process and thread features to emulate the process and thread facilities of its corresponding OS. An application client process issues its process creation request to the OS subsystem; then a process in the subsystem in turn issues a process request to the Windows executive. Windows enables the subsystem to specify the parent of the new process. The new process then inherits the parent's access token, quota limits, base priority, and default processor affinity.

1.1 Win32 API and Dynamic-Link Libraries

Windows programming [12] is useful to start off with some appreciation of some new terms intrinsic to Windows: objects, handles, instances, messages, and callback functions. These give us the mechanics of programming in this environment, that is, they are tools that we need to use. A Dynamic Link Library (DLL) is a module that contains functions and data that can be used by another module (application or DLL). DLLs provide a way to modularize applications so that their functionality can be updated and reused more easily. DLLs also help reduce memory overhead when several applications use the same functionality at the same time, because although each application receives its own copy of the DLL data, the applications share the DLL code. The Windows application programming interface (API) is implemented as a set of DLLs, so any process that uses the Windows API uses dynamic linking. Dynamic linking allows a module to include only the information needed to locate an exported DLL function at load time or run time.

Every process that loads the DLL maps it into its virtual address space. After the process loads the DLL into its virtual address, it can call the exported DLL functions. The system maintains a per-process reference count for each DLL. When a thread loads the DLL, the reference count is incremented by one. When the process terminates, or when the reference count becomes zero, the DLL is unloaded from the virtual address space of the process. Like any other function, an exported DLL function runs in the context of the thread that calls it. Therefore, the following conditions apply: 1) The threads of the process that called the DLL can use handles opened by a DLL function. Similarly, handles opened by any thread of the calling process can be used in the DLL function. 2) The DLL uses the stack of the calling thread and the virtual address space of the calling process. 3) The DLL allocates memory from the virtual address space of the calling process. Programs can be written in different programming languages can call the same DLL function as long as the programs follow the same calling convention that the function uses. The calling convention (such as C, C++, .Net and etc) controls the order in which the calling function must push the arguments onto the stack, whether the function or the calling function is responsible for cleaning up the stack, and whether any arguments are passed in registers. The Kernel32.dll is an important DLL for doing multithreaded programs; here the following important function we used in our research.

- Public Declare Function SetThreadAffinityMask Lib

```
"kernel32" Alias "SetThreadAffinityMask" (ByVal  
hThread As Long, ByVal dwThreadAffinityMask As  
Long) As Long
```

1.2 Symmetric Multiprocessing Support

Symmetric Multiprocessing (SMP) is a multiprocessing architecture in which multiple CPUs, residing in one cabinet, share the same memory. It is the tightly coupled process of program tasks being shared and executed, in true parallel mode, by multiple processors who all work on a program at the same time. Typically, these have large, single units with multiple processors that utilize shared memory, I/O resources, and, of course, a single operating system. The term SMP is so closely associated with shared memory that it is sometimes misinterpreted as standing for "shared memory parallel". SMP systems range from two to as many as 32 or more processors. However, if one CPU fails, the entire SMP system is down. Clusters of two or more SMP systems can be used to provide high availability (fault resilience). If one SMP system fails, the others continue to operate.

Windows supports an SMP hardware configuration. The threads of any process, including those of the executive, can run on any processor. In the absence of affinity restrictions, the microkernel assigns a ready thread to the next available processor. This assures that no processor is idle or is executing a lower-priority thread when a higher-priority thread is ready. Multiple threads from the same process can be executing simultaneously on multiple processors. As a default, the microkernel uses the policy of soft affinity in assigning threads to processors: The dispatcher tries to assign a ready thread to the same processor it last ran on. This helps reuse data still in that processor's memory caches from the previous execution of the thread. It is possible for an application to restrict its thread execution to certain processors (hard affinity).

Thread affinity forces a thread to run on a specific subset of processors. Setting thread affinity should generally be avoided, because it can interfere with the scheduler's ability to schedule threads effectively across processors. This can decrease the performance gains produced by parallel processing. An appropriate use of thread affinity is testing each processor. The system represents affinity with a bitmask called a processor affinity mask. The affinity mask is the size of the maximum number of processors in the system, with bits set to identify a subset of processors. Initially, the system determines the subset of processors in the mask. You can obtain the current thread affinity for all threads of the process by calling the GetProcessAffinityMask function. Use the SetProcessAffinityMask function [8] to specify thread affinity for all threads of the process. To set the thread affinity for a single thread, use the SetThreadAffinityMask function. The thread affinity must be a subset of the process affinity. On systems with more than 64 processors, the affinity mask initially represents processors in a single processor group. However, thread affinity can be set to a processor in a different group, which alters the affinity mask for the process.

2 IDENTIFYING PROBLEM

2.1 Multithreading in java

The Java run-time system [4] depends on threads for many things, and all the class libraries are designed with multithreading in mind. In fact, Java uses threads to enable the entire environment to be asynchronous. This helps reduce inefficiency by preventing the waste of CPU cycles. Java's multithreading system is built upon the Thread class, its methods, and its companion interface Runnable. Thread encapsulates a thread of execution. The Thread class defines several methods that help on managing threads.

After a computational job is designed and realized as a set of tasks, an optimal assignment of these tasks to the processing elements in a given architecture needs to be determined. This problem is called the scheduling problem [6] and is known to be one of the most challenging problems in parallel and distributed computing. The goal of scheduling is to determine an assignment of tasks to processing elements in order to optimize certain performance indexes. Performance and efficiency are two characteristics used to evaluate a scheduling system [9]. We should evaluate a scheduling system based on the quality of the produced task assignment (schedule) and the efficiency of the scheduling algorithm (scheduler). The produced schedule is judged based on the performance criterion to be optimized, while the scheduling algorithm is evaluated based on its time complexity.

In java, most of the executor [2] implementations in java.util.concurrent use thread pools. Thread pools address two different problems: they usually provide improved performance when executing large numbers of asynchronous tasks, due to reduced per-task invocation overhead, and they provide a means of bounding and managing the resources, including threads, consumed when executing a collection of tasks. Each ThreadPoolExecutor also maintains some basic statistics, such as the number of completed tasks. A ThreadPoolExecutor can additionally schedule commands to run after a given delay or to execute periodically. ScheduledThreadPoolExecutor class is preferable to Timer when multiple worker threads are needed, or when the additional flexibility or capabilities of ThreadPoolExecutor are required.

Java support flexible and easy use of threads; yet, java does not contain methods for thread affinity to the processors. Setting an affinity thread to multiprocessor [9] is not new to research, since it was already sustained by other multiprogramming languages for example C in UNIX platform [10] and C# in Windows platform [11]. This paper illustrates how java multithreaded program adapt with an affinity thread on multiprocessors in windows platforms.

3 PROBLEM SOLVING METHODS

3.1 Just Peculiar Algorithm (JPA)

Every Java application has a single instance of class Runtime that allows the application to interface with the environment in which the application is running. The current runtime can

be obtained from the getRuntime() method. This method returns the runtime object associated with the current Java application. Most of the methods of class Runtime are instance methods and must be invoked with respect to the current runtime object. The number of processors available to the Java virtual machine can be obtained through the availableProcessors() method. This value may change during a particular invocation of the virtual machine. Java applications are sensitive to the number of available processors should therefore occasionally poll this property and adjust their resource usage appropriately. Multithreaded programming is written in many programming languages with an improvement of setting an affinity to threads. However, Java does not contain any method to set an affinity for threads on CPU. Hence, we carried this research to synchronize all threads associated with the available processors. The following Program 1 is inherited Thread class to assign the affinity settings. The following is Just Peculiar Algorithm (JPA) algorithm shows how to set thread affinities by our own steps.

1. Get the available number of processor (Processors) in system
2. Select the processor number (ProcessorNum), where we assign thread to execute
3. Check whether selected processor number (ProcessorNum) is greater than and equal to the available processors (Processors) in the system or not. If selected processor is greater than and equal to available processor (Processors), throw IllegalArgumentException.
4. Initialize the incremental for loop with variable=0 and check the variable is less than available number of processor (Processors)
5. Check if the variable is equal to selected processor number (ProcessorNum) then create a thread.

The following Program 1 developed, based on the above algorithm with an affinity thread named as JThreadCore class and it packaged in javax.core, which can be exploit in user's applications. The JThreadCore class constructors and methods discussed bellow.

Constructors:

JThreadCore(String name)

JThreadCore(String name, int ProcessorNum).

The first constructor with the String name argument will assign the name of the thread. And the second constructor with String name argument will assign the name of the thread and int ProcessorNum will select the processor to execute the thread. To assign the processor, ultimately it calls the setAffinity(int ProcessorNum) to select the processor.

Methods:

synchronized void setAffinity(int ProcessorNum)

int getAffinity()

When two or more threads need to access to a shared processor, they need some way to ensure that the processor will be used by only one thread at a time. Consequently, the

setAffinity() method implemented with synchronized method, it can allow one thread can own a monitor at a given time. The argument of setAffinity(), int ProcessorNum will select the processor number to execute the Thread. And the getAffinity() method will return the affinity of the running thread.

Program 1: JThreadCore.java

```
// package name
package javax.core;
// inherited with Thread class
public class JThreadCore extends Thread{
// get the available system processor
private int Processors ;
// select the processor to execute the thread
private int ProcessorNum;
//assign the name of the thread
private static String name;
Thread t;
public JThreadCore(String name){
//interchange the thread name value to class variable
this.name=name;
}
public JThreadCore(String name,int ProcessorNum){
//interchange the thread name value to class variable
this.name=name;
//interchange the processor number to class variable
this.ProcessorNum=ProcessorNum;
//call setAffinity method
setAffinity(ProcessorNum);
}
//synchronized method to select the processor
public synchronized void setAffinity(int ProcessorNum){
// interchange processor number to class variable
this.ProcessorNum=ProcessorNum;
//get the available processor in the system
Processors = Runtime.getRuntime().availableProcessors();
//check selected processor is greater than equal to the selected
//processor
if (ProcessorNum>=Processors)
throw new IllegalArgumentException("This processor is not
available");
//create threads using loop
for(int i=0; i < Processors ; i++)
//check i value is equal to user selected processor
if (i==ProcessorNum){
//create thread
t=new Thread(name); }
}
//get the affinity of the thread
public int getAffinity(){
//return the core number, in which current thread is running
return ProcessorNum;
}
}
```

The following Program 2 developed to exercise the JThreadCore library. The Test class main main function has been created the object for Test("one",5,1,0) constructor. Once object created the constructor Test(String name,int wait,int

pri,int afi) automatically called. This assigns the name of thread as "one", waiting time as 5, priority as 1 and processor 1 to execute. The start() method in the constructor start the thread. Once thread started, it automatically call public void run() method, will execute until it reach the Thread.sleep() method. This Thread.sleep() will give chance, if other thread assigned on the same processor. Likewise, "two", "three" and "four" thread will execute on the assigned processor.

This program measured execution time using System.currentTimeMillis(). The formula to calculate the execution time is elapsedTime = (stopTime - startTime)/wait. Here startTime is starting time of the thread execution, stopTime is end of thread execution and wait is waiting time of each threads (same waiting time for all the threads). If suppose different waiting time assigned to thread, we should hold summation of all thread waiting time.

Program 2: Testing JThreadCore

```
import javax.core.JThreadCore;
class Test extends JThreadCore{
private static String name;
private static int wait;
private static int pri,afi;
Test(String name,int wait,int pri,int afi){
super(name);
this.name=name; this.wait=wait;
this.pri=pri; this.afi=afi;
setPriority(pri);
setAffinity(afi);
start();
}
public void run(){
try{
for(int i=0;i<5;i++){
System.out.println(i+"\t\t" + getId()+"\t\t\tCPU "+
getAffinity()+ "\t\t\t "+currentThread()); sleep(wait);
}
}catch(InterruptedException e){System.out.println(e); }
}
public static void main(String s[]){
long startTime,stopTime,elapsedTime;
try{
startTime = System.currentTimeMillis();
System.out.println("Value\tThread Id\tCPU #\t\t\tThread");
System.out.println("-----");
Test a1= new Test("one",5,1,0);
Test a2= new Test("two",5,1,1);
Test a3= new Test("three",5,10,0);
Test a4= new Test("four",5,10,1);
a1.join(); a2.join(); a3.join(); a4.join();
stopTime = System.currentTimeMillis();
elapsedTime = (stopTime - startTime)/wait;
System.out.println("Used execution time in ms: " +elapsedTime);
}catch(Exception e){System.out.println(e); }
}
}
```

This program is tested with with Inter Core™ 2 Duo CPU

T8100 @ 2.10 GHz (see Figure 1). Threads can run concurrently on assigned CPU absolutely, wherein threads "one" and "three" assigned on CPU 0, threads "two" and "four" assigned on CPU 2. The performance of this program evaluated on different multi-core environment by selecting different affinity on CPU, which is described on the Result and Discussion section. This research exercised with the available Java Thread library only to solve affinity thread problem. Since Java is a platform independent language, this program benefited to execute on any platform. The assigned thread names, not properly fixed on running threads are a big weakness of this research.

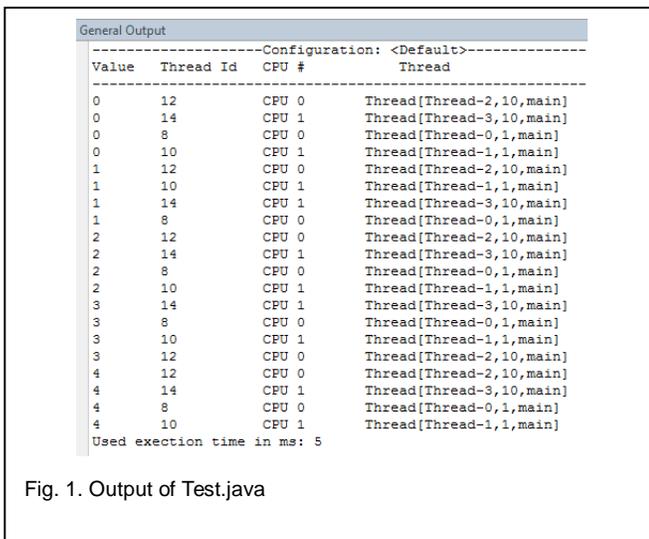


Fig. 1. Output of Test.java

3.2 Java Native Access (JNA)

Java Native Access provides [4] Java programs easy access to native shared libraries without using the Java Native Interface. JNA's design aims to provide native access in a natural way with a minimum of effort. The JNA library uses a small native library called foreign function interface library (libffi) to dynamically invoke native code. The JNA library uses native functions allowing code to load a library by name and retrieve a pointer to a function within that library, and uses libffi library to invoke it, all without static bindings, header files, or any compile phase. The developer uses a Java interface to describe functions and structures in the target native library. This makes it quite easy to take advantage of native platform features without incurring the high development overhead of configuring and building JNI code. JNA is built and tested on Mac OS X, Microsoft Windows, FreeBSD / OpenBSD, Solaris, and Linux. It is also possible to tweak and recompile the native build configurations to make it work on other platforms. For example, it is known to work on Windows Mobile, even if it is not tested for this platform by the development team.

If you've used the Java Native Interface (JNI) [7] to make a platform-specific native library accessible to your Java programs, you know how tedious it can be. Jeff Friesen continues his series on lesser-known open source Java projects by introducing you to Java Native Access -- a project that eliminates the tedium and error associated with JNI and lets

you access C libraries programmatically. In situations where Java does not provide the necessary APIs, it is sometimes necessary to use the Java Native Interface (JNI) to make platform-specific native libraries accessible to Java programs.

JNA approaches to integrate native libraries with Java programs. It shows how JNA enables Java code to call native functions without requiring glue code in another language. It is useful to know JNA, because the Java APIs with their architecture-neutral emphasis will never support platform-specific functionality. Though Java itself is architecture-neutral, JNA is perform on platform-specific. The Java Native Access project is hosted on Java.net, where you can download the project's online Javadoc and the software itself. Although the download section identifies five JAR files, you only need to download jna.jar. The jna.jar file provides the essential JNA software and is required to run all of the examples you'll find here. This JAR file contains several packages of classes, along with JNI-friendly native libraries for the Unix, Linux, Windows, and Mac OS X platforms. Each library is responsible for dispatching native method calls to native libraries. Here are a few things you have to take care of when starting a JNA project:

1. Download jna.jar from the JNA project site and add it to your project's build path. This file is the only JNA resource you need. Remember that jna.jar must also be included in the run-time classpath.
2. Find the names of the DLLs that your Java code will access. The DLL names are required to initialize JNA's linkage mechanisms.
3. Create Java interfaces to represent the DLLs such as kernel32.dll, user32.dll and etc on your application that will access.
4. Test linkage of your Java code to the native functions.

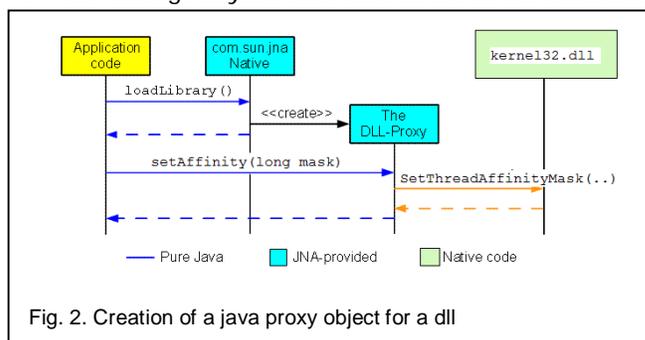


Fig. 2. Creation of a java proxy object for a dll

Many DLLs, such as those in the Windows API, host a large number of functions. But the proxy interface need only contain declarations for the methods your application actually uses.

3.2.1 A Proxy for the DLL

JNA uses the proxy pattern to hide the complexity of native code integration. It provides a factory method that Java programs use to obtain a proxy object for a DLL. The programs can then invoke the DLL's functions by calling corresponding methods of the proxy object. The code below shows an abbreviated view of a proxy interface for the Windows kernel32.dll.

Program 3. A proxy program for dll

```
//Kernel32.java
```

```
package javax.core;
import com.sun.jna.Library;
public interface Kernel32 extends Library {
    // Select the CPU
    int SetThreadAffinityMask(int threadid,int mask);
    int GetCurrentThreadId(); // Get thread Id
    int Sleep(long Milliseconds); // Assign waiting time
}
```

JNA takes care of all run-time aspects, but it requires your help to create the proxy's Java class. So the first piece of code you need to create is a Java interface with method definitions that match the DLL's C functions. To play with JNA's run-time correctly, the interface must extend com.sun.jna.Library.

3.2.2 Linkage of java code to the native functions

The following KThreadCore.java uses the User32 interface shown above to create a proxy for the Windows Kernal32.DLL. When setAffinity(long mask) method is executed, which in turn invokes the DLL's SetThreadAffinityMask() function. The run-time mapping of the proxy method to the DLL function is handled transparently by JNA, the user just has to ensure that the method name matches the function name exactly.

Program 4. Linkage of java code to the native function

```
// KThreadCore.java
package javax.core;
import com.sun.jna.Native;
import javax.core.Kernel32;
public class KThreadCore extends Thread {
    Kernel32 kernel321=(Kernel32)
    Native.loadLibrary("kernel32", Kernel32.class);
    int mask;
public KThreadCore(){ }
public KThreadCore(String thread_Name){
    super(thread_Name);
}
public int setAffinity(int tid,int mask) {
int mask1= Runtime.getRuntime().availableProcessors();
if (mask>mask1)
throw new IllegalArgumentException("The CPU mask should
starts from 1..N");
this.mask=mask;
return kernel321.SetThreadAffinityMask(tid,mask);
}
public int getAffinity(){
    return mask;
}
public int getCurrentThreadId(){
    return kernel321.GetCurrentThreadId();
}
public int context(int duration){
    return kernel321.Sleep(duration);
}
}
```

In Program 4, KThreadCore class extends with Thread class consequently KThreadCore class will inherit all the methods of Thread class, which will be used in Program 3. The kernel321 object is a loaded with class library that includes the declared

function of Kernel32.java (Program 3). The KThreadCore() constructor will create threads without name of the thread and KThreadCore(String thread_Name) will create threads with specified of the threads using the constructor argument. The setAffinity(long tid,long mask) is alias method for SetThreadAffinityMask(long threadid,long mask) on proxy DLL. The setAffinity(long tid,long mask) method has an exception mechanism to check whether the specified CPU is available or not. If CPU is present then it will call the proxy dll SetThreadAffinityMask(long threadid,long mask) to assign the affinity for a particular thread. If CPU is not present then it will convey the exception with "The system has + mask1 + CPU only and the specified CPU + mask + is not available in the system". The getAffinity() method will return the affinity of the thread. The getCurrentThreadId() call the proxy DLL's GetCurrentThreadId(), which will assign thread id to every threads. The context(int duration) call the proxy DLL's Sleep(long Milliseconds) method, which will assign the thread to wait for specified milliseconds to avail other threads to execute on CPU.

3.2.3 Application code

Until now, we created the proxy DLL (Program 3) and linkage of java code to the native function (Program 4). This section shows you to test the thread with an affinity on CPUs.

Program 5. Testing the affinity thread

```
// Test.java
import javax.core.KThreadCore;
class Test extends KThreadCore{
String name;
static int wait,affinity;
int pri;
Test(String name,int wait,int pri,int affinity) throws Exception{
    super(name);
    this.name=name;
    this.wait=wait;
    this.pri=pri;
    this.affinity=affinity;
    setPriority(pri);
    setAffinity(getCurrentThreadId(),affinity);
    start();
}
public synchronized void run(){
try{
for(int i=0;i<5;i++){
System.out.println(i+"\t\t"+getCurrentThreadId()
+"\t\tCPU " +getAffinity()+"\t\t" +currentThread());
context(wait);
}
}catch(Exception e){System.out.println("Error :"+e);}
}
public static void main(String aa[]){
long startTime,stopTime,elapsedTime;
try{
startTime = System.currentTimeMillis();
System.out.println("Value\tThread Id\tCPU #\t\t\tThread");
System.out.println("-----");
```

```
//(thread name, waiting time, priority, affinity)
Test a1=new Test("ONE",5,1,0);
Test a2=new Test("TWO",5,1,1);
Test a3=new Test("THREE",5,10,0);
Test a4=new Test("FOUR",5,10,1);
// joins waiting thread through Thread class method
a1.join();      a2.join();      a3.join();      a4.join();
stopTime = System.currentTimeMillis();
elapsedTime = (stopTime - startTime)/wait;
System.out.println("Used execution time in ms: " +elapsedTime);
}catch(Exception e){System.out.println("Error : " + e);}
}
```

In Program 5, Test class extends with KThreadCore class consequently all Thread class methods can be accessed. In the main function an object is created for Test(String name,long wait,int pri,long affinity) constructor. This constructor has the argument for name of the thread, waiting time, priority, and affinity for a thread. Once an object is created, a constructor is automatically called. In the constructor super(name) method, the thread is created by passing the "name" to the KThreadCore(String thread_Name) constructor. The setPriority(pri) is actually from Thread class, which will set the priority for a thread and setAffinity((getThreadId(),affinity) is our method from KThreadCore class, which will set the affinity for a thread, here getThreadId() method get current thread id for a thread to set affinity on CPU. Once start() method starts the thread, it will automatically call the public void run() method. The run() method contain a loop to print the value from 0 up to 5. Test class created with four objects a1, a2, a3 and a4 respectively as ONE, TWO, THREE and FOUR threads. The ONE and TWO threads have 1 (MIN_PRIORITY) priorities and the affinities of threads are 0 and 1 respectively; The THREE and FOUR threads have 10 (MAX_PRIORITY) priorities and the affinities of threads are 0 and 1 respectively. Hence, ONE and THREE threads executed on processor CPU 0 with respect to priority; similarly, TWO and FOUR threads are expected to execute on processor CPU 1 with respect to priority. This program is tested with with Inter Core™ 2 Duo CPU T8100 @ 2.10 GHz each (see Figure 3).

General Output			
Value	Thread Id	CPU #	Thread
0	4968	CPU 0	Thread[ONE, 1, main]
0	5928	CPU 0	Thread[THREE, 10, main]
0	3996	CPU 1	Thread[FOUR, 10, main]
0	4964	CPU 1	Thread[TWO, 1, main]
1	5928	CPU 0	Thread[THREE, 10, main]
1	4964	CPU 1	Thread[TWO, 1, main]
1	4968	CPU 0	Thread[ONE, 1, main]
1	3996	CPU 1	Thread[FOUR, 10, main]
2	3996	CPU 1	Thread[FOUR, 10, main]
2	5928	CPU 0	Thread[THREE, 10, main]
2	4964	CPU 1	Thread[TWO, 1, main]
2	4968	CPU 0	Thread[ONE, 1, main]
3	3996	CPU 1	Thread[FOUR, 10, main]
3	5928	CPU 0	Thread[THREE, 10, main]
3	4964	CPU 1	Thread[TWO, 1, main]
3	4968	CPU 0	Thread[ONE, 1, main]
4	5928	CPU 0	Thread[THREE, 10, main]

4 RESULTS AND DISCUSSION

4.1 Thread Migration

Thread migration is when threads in computer core are able to move from core to another core. Just Peculiar Algorithm (JPA) enables to set the affinity to the thread on multi-core systems. This research finding will enables threads migration in ready, waiting and running states of threads. The following fragmented code added in Program 2 shows, how a thread can migrate between cores.

```
for(int i=0;i<5;i++)
if(i==2 and getID()==14)
setAffinity(1);
.....
2 14 CPU 0 Thread[Thread-3,10,main]
2 10 CPU 0 Thread[Thread-1,10,main]
2 8 CPU 0 Thread[Thread-0,1,main]
3 14 CPU 1 Thread[Thread-3,10,main]
.....
```

Here, when a threadID 14 reaches the loop iteration 2, then it migrate the thread from core 0 to core 1. This is done for a single thread. But when more number of threads needed to be migrate, then like other platform thread should have a queue. For example, Linux thread migration mechanism[1], normally used for relatively long-term load-balancing across cores. To our knowledge, Linux thread migration mechanism is the current state of the art for core-switching. When a task wants to migrate, it puts itself on a per-core migration queue, wakes up and switches control to a per core migration thread, which does the actual work of moving the thread to the run queue of the target core.

4.2 The Performance issue

The performance of Just Peculiar Algorithm (JPA) (described in section 3.1) and Java Native Access (JNA) (described in section 3.2) evaluated on different multi-core environment by selecting different affinity on CPU and adding few more threads with different iterations.

TABLE 1
 EXECUTION TIME IN MILLISECONDS

Methods	Core 2 Duo				Core i5				Core i7				
	Iterations												
	5	25	50	100	5	25	50	100	5	25	50	100	
Four threads	JPA	5	25	96	261	5	26	51	102	5	29	62	107
		5	25	97	265	5	27	52	101	5	29	59	102
	JNA	5	25	102	260	5	26	51	102	5	31	61	103
		5	26	101	262	5	26	51	102	5	28	62	105
Seven threads	JPA	26	44	94	248	29	54	78	126	31	55	79	130
		25	47	90	242	32	53	76	127	32	56	88	132
		27	46	92	246	31	50	77	130	33	57	91	133
		28	53	95	267	28	51	74	128	34	64	81	129
	JNA	5	79	222	504	3	25	49	100	5	48	51	101
		7	82	225	510	4	28	50	104	5	50	99	109
		5	78	229	507	5	30	53	106	4	51	102	103
		5	81	218	503	5	29	55	103	5	52	52	105
JNA	26	83	207	476	43	106	184	343	43	106	181	343	
	27	73	210	471	46	109	187	340	46	109	184	340	
	37	81	212	494	43	107	190	343	43	107	187	343	
	36	95	213	493	46	108	185	340	46	108	190	340	

We analyze about the execution on different computer core with the time in milliseconds, which is described in the above Table 1. Just Peculiar Algorithm (JPA) is just a techniques to set the affinity through Java Thread library, hence the execution speed is somewhat little than Java Native Access (JNA). Because, the Java Native Access (JNA) is approaching windows DLL to perform the affinity schedule on the processor, hence it acquired more time than Just Peculiar Algorithm (JPA).

5 CONCLUSION

Working with multiple threads on symmetric multiprocessor is very natural to improve the performance based on number of CPUs. Thread affinity benefit a thread to run on a specific subset of processors, enable us to schedule threads on a particular CPU. Setting an affinity to CPU is not new to research. Since affinity thread is already sustained by other multiprogramming languages on different platforms. Java does not have method to set affinity for a thread. While Java exercised already with affinity threads in UNIX platforms through Java Native Interface. There is lack in windows platform for Java affinity thread. This paper exemplified the method to set an affinity for threads to execute on particular CPU through Java Java Peculiar Algorithm (JPA) and Native Access (JNA) in windows platforms. The performance metric also deliberated.

REFERENCES

1. ACM SIGOPS Operating Systems Review special issue on The Interaction Among the OS, the Compiler, and Multicore Processors, April, 2009, Volume 43, Number 2.
2. API reference, <http://docs.oracle.com/javase/7/docs/api/index.html>
3. Brian goetz, tim peierls, joshua bloch, joseph bowbeer, david holmes, doug lea, java concurrency in practice, addison wesley professional, may 09, 2006, isbn-10: 0-321-34960-1
4. Get started with jna, <https://jna.dev.java.net>
5. Herbert schildt, java™ 2: the complete reference, fifth edition, mcgraw-hill, 2002, isbn:0-07-222420-7
6. Hesham el-rewini, mostafa abd-el-barr, 2005, advanced computer architecture and parallel, a john wiley & sons, inc publication.
7. Java Native Interface, <http://java.sun.com/javase/6/docs/technotes/guides/jni/index.html>
8. Microsoft developer network (msdn): setprocessaffinitymask <http://msdn.microsoft.com/library/default.asp?url=/library/enus/dllproc/base/setprocessaffinitymask.asp>
9. Mike loukides, gian-paolo d. Musumeci, system performance tuning, second edition, o'reilly, 2002, isbn : 0-596-00284-x
10. parallel programming? Well, it's all about cpu affinity or how to set processor affinity in wpf <http://khason.net/blog/parallel-programming-well-it%e2%80%99s-all-about-cpu-affinity-or-how-to-set-processor-affinity-in-wpf/>
11. Take charge of processor affinity, <http://www.ibm.com/developerworks/linux/library/l-affinity/index.html>
12. The win32 programming tutorials for fun (and funny too), <http://www.installsetupconfig.com/win32programming/>

Beamforming in Wireless Sensor Networks

Jasmine K.S, Sumithradevi K.A

Abstract— Wireless interconnection is now considered a defining feature for sensor networks. The sensors are visualized as miniature battery-driven computing devices; because of energy constraints, all computing, storage/retrieval, and especially communication operations are considered expensive. Scalability and robustness (to sensor failures) are considered essential features for a sensor network. The limitations of sensor devices is that they have small batteries with small amounts of available power; thus limiting the lifespan of the sensors. An attractive solution to making the wireless sensors more energy efficient is to use cooperative beamforming. Implementing cooperative beamforming clearly comes with power and time overhead for the data sharing among the collaborative sensors. Implementing cooperative beamforming is a multi-variable optimization problem. In this paper various beamforming techniques along with a special emphasis on cooperative beamforming in acoustics with the help of Antenna technology is discussed which can save energy over sensor transmissions.

Index Terms— Beamforming, Sensor Networks, Wireless Sensors, Antenna Technology, Cooperative beamforming, Collaborative sensors, Distributed MIMO.



1 INTRODUCTION

There is an ever-increasing demand on wireless communication servers to provide voice and high-speed data services. At the same time, to support more users per basestation to reduce overall network costs and make the services affordable to its users. As a result, wireless systems that enable higher data rates and higher capacities are a pressing need. Unfortunately, because the available broadcast spectrum is limited, attempts to increase traffic within a fixed bandwidth create more interference in the system and degrade the signal quality. Beamforming is a ubiquitous technology when it comes to wave propagation. Beamforming is a signal processing technique used in sensor arrays for directional signal transmission or reception. This is achieved by combining elements in the array in a way where signals at particular angles experience constructive interference and while others experience destructive interference. Beamforming can be used at both the transmitting and receiving ends in order to achieve spatial selectivity. The improvement compared with an omnidirectional reception/transmission is known as the receive/transmit gain (or loss).

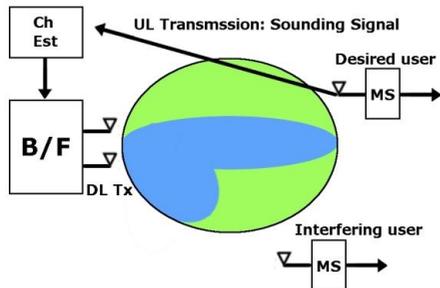


Fig. 1. Beamforming

2 ANTENNA TECHNOLOGY AND BEAMFORMING

Sensor network communication systems suffer from interference problem (intra-cell, inter-cell). Thus, the interference issue is a key factor in designing the system. Many techniques were developed and now used for mitigating the interference problem. These techniques attempt to reduce the interference effects on the performance through increasing SIR and user capacity. Smart antennas are an interference reduction method. With the use of this technology, two types of gain can be achieved, multiplexing gain which leads to higher data rates, and diversity gain which leads to better reliability. Smart antennas are compatible with Multiple Input Multiple Output (MIMO) systems [11]. Smart antenna technology offers a significantly improved solution to reduce interference levels and improve the system capacity. With this technology, each user's signal is transmitted and received by the basestation only in the direction of that particular user. This drastically reduces the overall interference in the system. A smart antenna system, as shown in Figure 2, consists of an array of antennas that together direct different transmission/reception beams toward each user in the system.

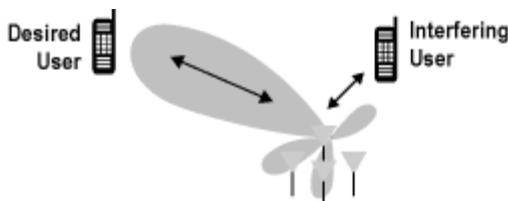


Fig. 2. Smart Antenna System and Beamforming

In beamforming, each user's signal is multiplied with complex weights that adjust the magnitude and phase of the signal to and from each antenna. This causes the out-

- Jasmine K.S, Dept. of MCA, RVCE, India, PH-080-67178159, E-mail: jkjasmine1@gmail.com
- Sumithraka, Dept. of MCA, RVCE, India, PH-080-67178159, E-mail: sumithraka@gmail.com

put from the array of antennas to form a transmit/receive beam in the desired direction and minimizes the output in other directions.

Beamforming has found numerous applications in radar, sonar, seismology, wireless communications, radio astronomy, speech etc. Because of the wide range of applications of beamforming, there exist multiple design criteria for beampatterns. They include creating a beampattern that matches a desired beampattern, having a minimum beamwidth for a certain sidelobe level or null steering.

The solution to these different beampattern designs is almost as numerous as the number of applications of beamforming [12]. Some examples are:

- Minimum Variance Distortionless Response beamforming
- Statistical Eigen beamforming
- Beamspace beamforming
- Frost beamforming
- Generalized Sidelobe Cancellers

2.1 Switched and Adaptive Beamforming

Beamformers can be classified as either data independent or statistically optimum, depending on how the weights are chosen. The weights in a data independent beamformer do not depend on the array data and are chosen to present a specified response for all signal / interference scenarios. The weights in a statistically optimum beamformer are chosen based on the statistics of the array data to optimize the array response. In general, the statistically optimum beamformer places nulls in the directions of interesting sources in an attempt to maximize the signal to noise ratio at the beamformer output. Data independent beamformer design techniques are often used in statistically optimum beamforming. The statistics of the array data are not usually known and may change over time so adaptive algorithms are typically employed to determine the weights.

If the complex weights are selected from a library of weights that form beams in specific, predetermined directions, the process is called switched beamforming. Here, the basestation basically switches between the different beams based on the received signal strength measurements. On the other hand, if the weights are computed and adaptively updated in real time, the process is called adaptive beamforming. Through adaptive beamforming, the basestation can form narrower beams towards the desired user and nulls towards interfering users, considerably improving the signal-to-interference-plus-noise ratio.

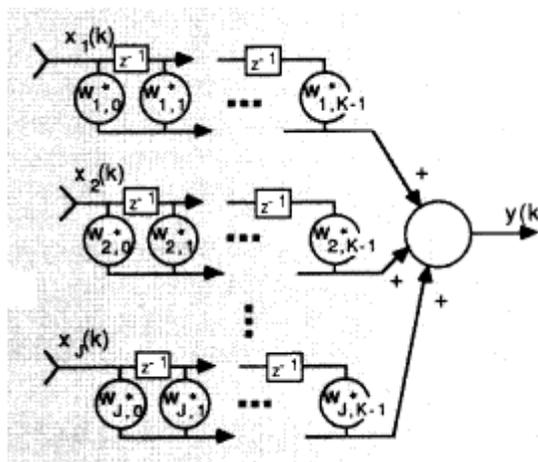


Fig. 3. Common broadband beamformer

2.2 Acoustic Beamforming

Finding out the exact source of a sound is a tough challenge for any acoustics engineer. A number of methods, based on microphone arrays are into practice. In general, the methods fall into three categories: near-field acoustic holography, acoustic beamforming, and inverse methods[4,5]. Depending on the test object, the nature of the sound and the actual environment, engineers will have to select one method or the other.

These are two important criteria to assess the validity of sound source localization methods:

- **Spatial resolution** is the ability to separate two sound sources. This is usually expressed in centimeters. It represents the closest distance between two sources, where they still appear to separately and do not merge into a single source. The lower the spatial resolution, the better the source localization.
- **Dynamic range** expresses sound level differences in dB between real sound sources and their surrounding mathematical artifacts inherent to the sound source localization techniques. The higher the dynamic range, the better the source localization.

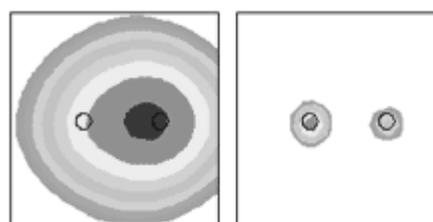


Fig. 4. Spatial Resolution

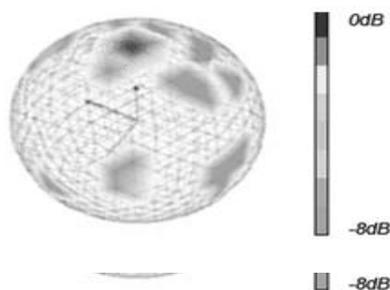


Fig. 5. Dynamic range

Acoustic beamforming is a technique where the microphone array is placed in the far field. As a rule of thumb, the far field is defined as being further away from the source than the array dimensions or diameter. The area between near field and far field remains a grey zone. In the near field, sound waves behave like circular or spherical waves whereas, in the far field, they become planar waves.

Numerous microphone configurations are possible in acoustic beamforming arrays. In general, the configuration is usually a trade-off between dynamic range and source localization accuracy. To get the best of both worlds, it is preferred to select a circular array with a pseudo-random microphone distribution.

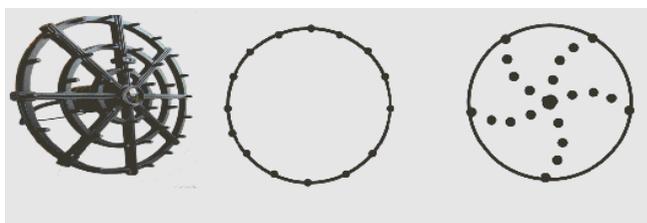


Fig. 6. Array configurations: full pseudo random (left), ring only (mid), spiral (right)

The ring array in middle of figure 2 provides good results when the exact distance to the source is not known, but dynamic range is low. Also, spiral-shaped arrays on the right result in lower dynamic range, compared to arrays that have more microphones distributed over the entire area of the array. More important, arrays without uniform microphone distribution will not have enough dynamic range when used in the near field.

The acoustic beamforming technique was first developed for submarines and environmental applications. In the far field, sound waves hitting the array are planar waves. Under these conditions, it is possible to propagate the measured sound field directly to the test object. All microphone signals measured by the acoustic beamforming array are added together, taking into account the delay corresponding to the propagation distance. The pressure

can be calculated at any point in front of the array, allowing propagation to any kind of surface. Acoustic beamforming is sometimes called “sum and delay” since it considers the relative delay of sound waves reaching different microphone positions. Acoustic beamforming requires that all data is measured simultaneously.

2.2.1 Advantages and disadvantages of acoustic beamforming

Acoustic beamforming has the following advantages:

- Propagation does not relate to the size of the measurement array. The test object can be larger than the array. Since all data is measured simultaneously, results can be viewed almost instantly after data acquisition.

Because of the relatively fast acquisition and analysis speed, acoustic beamforming lets engineers evaluate several configurations in a limited amount of time.

This flexibility has some negative aspects:

- The spatial resolution is proportional to the wavelength:

$$spatial\ resolution = \frac{d}{D} \lambda$$

Where d is the distance between the source and array, D the array diameter, and λ the wavelength. In an ideal situation, when the antenna is at a distance D to the source, the resolution is equal to the wavelength. If the array is placed farther from the structure, the resolution becomes worse. Acoustic beamforming, in general, is only usable at frequencies above 1000Hz.

- Acoustic beamforming can not be used to calculate sound power. Proper source ranking cannot be done with this technique.

2.2.2 How can the disadvantages be overcome

The main disadvantage is that acoustic beamforming does not perform well in the low frequency range. This can be improved by using a dedicated acoustic beamforming technique called *near-field focalization*.

Near-field focalization is a beamforming technique that uses measurements in the near field, whereas classical acoustic beamforming is measured in the far field. In the near field, the sound waves no longer arrive at the microphone as planar waves, but as spherical waves. The original beamforming back propagation is reformulated to deal with these waves. Near-field focalization improves that spatial resolution to 0.44 .

2.2.3 How can acoustic beamforming and near-field acoustic holography be combined?

While acoustic beamforming provides the best results when using an array with a pseudo-random microphone distribution, near-field acoustic holography (NAH) requires a rectangular array with evenly spaced microphones - both horizontally and vertically. Horizontal and vertical spacing can be different. NAH can not be performed with an acoustic beamforming array as (1) it is not rectangular and (2) it has a pseudo-random microphone distribution.

To overcome this, the problem is rewritten as an inverse method. The transfer function in this formulation includes both propagative and so-called evanescent wave functions, and needs an optimal and stable principle component analysis-based regularization which includes evanescent wave filtering. The method is called irregular near-field acoustic holography or irregular-NAH.

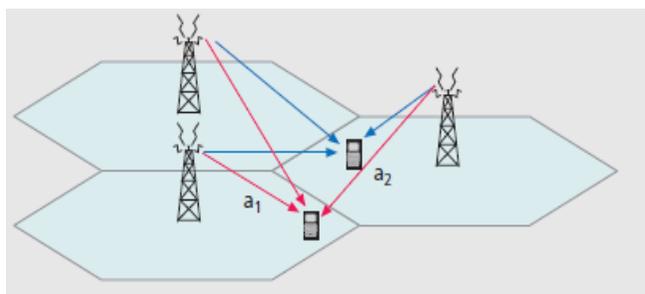


Fig. 7. COMP (cooperative MIMO)

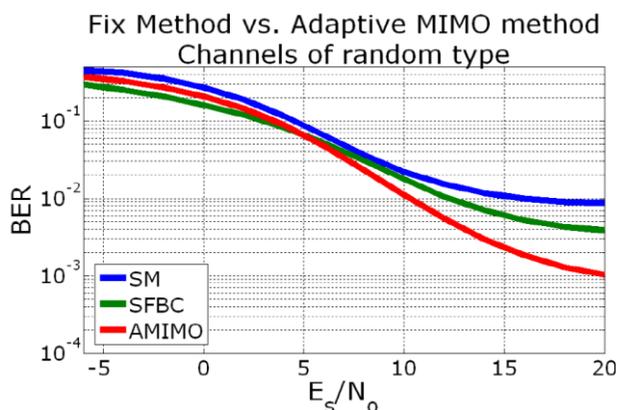


Fig. 8. Fixed Vs. Adaptive MIMO

2.3 Cooperative MIMO

Multiple-input multiple-output (MIMO) is an advanced

technology that can effectively exploit the spatial domain of mobile fading channels to bring significant performance improvements to wireless communication systems. Conventional MIMO systems, known as point-to-point MIMO or collocated MIMO, require both the transmitter and receiver of a communication link to be equipped with multiple antennas. In practice, however, many wireless devices may not be able to support multiple antennas due to size, cost, and/or hardware limitations. Cooperative MIMO [14], also known as virtual or distributed MIMO, aims to utilize distributed antennas on multiple radio devices to achieve some benefits similar to those provided by conventional MIMO systems.

The basic idea of cooperative MIMO is to group multiple devices into virtual antenna arrays (VAAs) to emulate MIMO communications. A cooperative MIMO transmission involves multiple point-to-point radio links, including links within a VAA and links between possibly different VAAs. For relay-based cooperative MIMO communications, there are three main cooperative strategies: amplify-and-forward, decode-and forward, and compress-and-forward techniques.

Previous theoretical studies have revealed the pros and cons of cooperative MIMO compared to point-to-point MIMO systems [1]. The disadvantages of cooperative MIMO come from the increased system complexity and the large signaling overhead required for supporting device cooperation. The advantages of cooperative MIMO, on the other hand, are due to its capability to improve the capacity, cell edge throughput, coverage, and group mobility of a wireless network in a cost-effective manner. These advantages hinge on the usage of distributed antennas, which can increase the system capacity by decorrelating the MIMO channels and allow the system to exploit the benefits of macro-diversity in addition to micro-diversity. In many practical applications, such as cellular mobile and wireless ad hoc networks, the advantages of deploying cooperative MIMO technology usually outweigh the disadvantages [13].

3. CONCLUSION

As a general rule, near-field techniques should be preferred for sound source localization. They provide the best results in terms of dynamic range and spatial resolution. There are situations where a near-field technique is not applicable: (1) it is not possible to measure in the near field, (2) the array size becomes too big, or (3) it is not possible to measure in patches due to rapidly changing operational conditions. In these cases, an acoustic beamforming solution will be chosen.

Acoustic beamforming with near-field focalization is a good alternative, providing results with good spatial resolution and dynamic range, depending on the frequency

range. It uses an array with pseudo-random distributed microphones. Acoustic beamforming obtains analysis results in a single shot wide-angle measurement, making it an ideal tool for troubleshooting, as it offers a quick preview with improved spatial resolution using near field focalization, but also for in-depth root cause analysis, when using the irregular-NAH technique.

Distributed transmit beamforming is a form of cooperative communication in which two or more information sources simultaneously transmit a common message and control the phase of their transmissions so that the signals constructively combine at an intended destination. Depending on the design objectives and constraints, the power gains of distributed beamforming can be translated into dramatic increases in range, rate, or energy efficiency. Distributed beamforming may also provide benefits in terms of security and interference reduction since less transmit power is scattered in unintended directions. Key challenges in realizing these benefits, however, include coordinating the sources for information sharing and timing synchronization and, most crucially, distributed carrier synchronization so that the transmissions combine constructively at the destination.

REFERENCES

- [1] J.D. Maynard, E.G. Williams and Y. Lee, "Near-field acoustic holography I. Theory of generalized holography and the development of NAH", *The Journal Acoustical Society of America.*, October 1985, Volume 78, Issue 4, pp 1395-1413
- [2] E.G. Williams, "Fourier Acoustics - Sound Radiation and Near-field Acoustical Holography", Academic Press, 1999.
- [3] Noise and Vibration, Near-field Acoustical Holography, "Noise Source Identification and Quantification", www.lmsintl.com
- [4] Finn Jacobsen and Yang Liu, Near-field acoustic holography with particle velocity transducers, *The Journal Acoustical Society of America*, November 2005, Volume 118, Issue 5, pp 3139-3144.
- [5] J. Hald, "Time Domain Acoustical Holography and Its Applications", *Sound &Vibration*, B&K Technical Review, 16-25, February 2001.
- [6] J. Hald, "Beam-forming and Wavenumber processing", *B&K Technical Review*, No. 1, 2004.
- [7] J. Hald, "Combined NAH and Beam-forming Using the Same Microphone Array", *Sound & Vibration*, 18-27, December 2004
- [8]Nagle, D., J. Edelson and L. Freitag , "Medium-Frequency Tactical Acoustic Communication System Design and Performance (U), U. S", *Navy Journal of Underwater Acoustics*, 50(4): pp. 961-987, Oct. 2000
- [9]Kilfoyle, D. J. Preisig, and A. Baggeroer., "Spatial Modulation over Partially Coherent Multiple-Input / Multiple-Output Channels", *IEEE Transactions on Signal Processing* , 51(3): p. 794-804, 2003.
- [10] Kilfoyle, D. and L. Freitag, "Spatial Modulation in the Underwater Acoustic Channel", *Proc. Array, Sensors, and Adaptive Proc. Workshop*, Lincoln Laboratories, Mar. 2004.
- [11]Y. Fan and J. S. Thompson, "MIMO Configurations for Relay Channels: Theory and Practice," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, May 2007, pp. 1774-86.
- [12] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network Coordination for Spectrally Efficient Communications in Cellular Systems," *IEEE Commun. Mag.*, vol.44, no. 8, Aug. 2006, pp. 56-61.
- [13] H. Bolcskei, D. Gesbert, C. Papadias, and A. van der Veen, *Space-Time Wireless Systems*. Cambridge University Press, 2006.
- [14] Huaiyu Dai , Hongyuan Zhang and Quan Zhou, "Some Analysis in distributed MIMO Systems", *Journal Of communications*, Vol. 2, No. 3, May 2007.

Law and Economics: Environmental Costs and Benefits

Clifford D. Fisher

Abstract— In an increasingly litigious society the courts are forced to entertain a myriad of topics. One of the most muddled and polluted involves environmental law. When our legislative branch of government is hesitant to act against polluters on their own free will, environmentalists and victims turn to the courts to resolve their problems. The courts are attempting to meet this influx with the principles and guidelines that have sufficed for centuries in tort law. Why will this same process not satisfy environmental litigation today? A combination of politicians, the public, businesses, economists, and scientists pollute this tradition, in a web of technicalities, complications in causation, and unknown answers.

Politicians are the puppets of society's playhouse. They respond when their strings are pulled. They are torn by their loyalty to the businesses, who financially back them, and by those who reelect them. Corporations pull their arms in the direction of less stringent regulations and less power among agencies, while citizens jerk their legs in the direction of protection of public risks. But this protection comes at what cost? If a casual link exists between anything, it is the public's blind hysteria to something the media claims to cause cancer, and a politician plowing money into the newest fund to abate the problem...often supporting a ban on the basis of fear, not knowledge. Could the money be spent more wisely? Usually if dealing with the government, the resounding answer is "YES"!

Business and economists base funding on more cost-effective terms. In an attempt to stifle and constrict the powers of the EPA, President Reagan propelled an economic movement in the reduction of abatements with Executive Order 12291. This was a significant introduction of cost-benefit analysis (CBA) into the political mainstream. In accordance with Pareto Superior and the theories of Coase and Pigot, CBA is a strategy that indicates the most cost-effective policies for achieving regulatory control. This is accomplished by aiding in determining if a proposal for development would be more beneficial to society than harmful. Through this process, economists are obligated to endorse those products or projects that yield a higher return. Although tenable, this analysis is not without its faults. CBA is a function of evaluating opportunity costs and placing a value on intangible and indispensable items, such as life.

One of the concerns of scientists/environmentalists is that benefits, from their standpoint, are not easily calculated. It is very difficult to determine whose life has greater value. According to a lawyer or businessperson, it may be calculated utilizing potential income lost, but an ecologist would argue that every life is priceless. Whether human life or clean air and water, to give a monetary value is to consider it as a commodity to be bought and sold in the market. Approaches to controlling the costs and benefits of environmental control vary as they are viewed through different disciplines. As for the public opinion, how much is your life worth?

Science is concerned with objective truths and theories that can be verified by those who have extensive training in scientific method. On the other hand, law involves rules and regulations translated by judges and juries, who lack the knowledge to adequately and thoroughly evaluate scientific evidence. The law allows and depends upon the testimony and information provided by experts. However, there is some discrepancy as to how responsive the law is to science and science to the law. Often the legal decision must be made before the evidence is definitive.

Current environmental issues on trial require a far deeper erudition of chemical reactions, the function of ecosystems, and the limit of safe exposure to carcinogens, which outreaches the scope of knowledge possessed by traditional lawyers. To properly combat the antagonist of environmental degradation, specialization in all fields of the environment would be necessary. As an old factory would have to update their production methods to meet new environmental standards (often at a great cost) practicing lawyers would have to update their array of knowledge to also meet new environmental standards for law.

The process and the need to assign and determine a cause vary greatly in the fields of science and law. Law focuses on fairness. In the interest of time, those who do not have all of the conclusive evidence must make a decision. In comparison, scientists seek but rarely find "absolute truth". Some of the most difficult and costly issues in health and the environment are created by uncertain causation occurring in tort litigation and environmental cleanup. Some diseases or environmental hazards do not possess a single cause, but a constellation of component causes. Some effects are immediate while most are latent. In these instances, the reason of detriment cannot be determined by sound science alone. Issues of liability will arise and costly court cases and/or costly settlements will result. Such examples include environmental cleanup of water where the sources of contamination are uncertain, or in compensation awards for such ailments as cancer. Jumping to the conclusion that there is a causal link between cancer and chlorine may do more damage than good. This is true especially if those funds can be allocated to environmental cleanup or research in other areas.

Environmental law falls under a legal category shaded gray. There are no definitive answers, concurring universal conclusions, or well-established standards. However, there is imperfect information. Common law is not equipped to handle damages yet unknown or cases where a multitude of polluters are involved. The courts are not equipped to make decisions without the help of scientists, but they speak two different languages. Untrained lawyers are unfit to argue an adequate defense in ignorance to the medical and scientific terms and logic. Juries are unable to render a just verdict with little or no knowledge on how to interpret the data that contains an admission. So where does that leave an environmentally conscious citizen, harmed victim, or potentially endangered family with little hope in this confusion among professionals, and those entrusted to protect and defend?



1 INTRODUCTION

IN trying to determine the value of life, or the costs and benefits associated with it, law and economics make certain basic assumptions, and with these assumptions they do some things well and others poorly. Because of its concentration on individual conduct, economics tends not to see or accommodate concerns about culture, community, or nature. When determining causation, economic perspective makes it seem like actions are the result of individuals and not the government.

One central issue in all of this is that disciplines such as economics, biology, geology, sociology, philosophy, or business, are now starting to be used as normative guides for conduct by government. In other words, there is no longer just economics or just science, but economics as law or science as law. With this new perspective, the usual, safe assumptions are open to public debate. In order to understand this debate, it is essential to gain an understanding of environmental issues and the strengths and weaknesses associated with different risks. Only then will we be able to wisely determine how to improve our environmental law system.

2 CONTRASTING OUTLOOKS: PUBLIC RISKS VS. PRIVATE RISKS

In dealing with environmental risks, questions arise regarding what risks should be targeted. Should we target risks that affect the most people? Should we target risks that may not affect many people, but that have very severe consequences for those affected? A problem is, though, that peoples' perceptions of risk are all different. Depending on their backgrounds, education levels, places of residence, etc., people perceive risks differently. Plus, with some environmental issues, because of long latency periods or unknown effects, people cannot possibly assess risks accurately. Even if risks could be properly assessed and people agreed upon which risks are most threatening, yet another question arises: Do we even have the right to regulate these risks for people? Upon each of these questions lies an array of different opinions stemming from different perspectives.

The Environmental Protection Agency (EPA) and Congress are responsible for protecting large populations from major risks to health and safety which are centrally produced, often in large quantities, and which reside largely outside the personal understanding or control of individuals [1]. In determining what constitutes societal benefits of environmental regulation, scientists and government have focused on risks associated with life. This concept deals with risk assessment because if there was no possibility of injury to individuals, there would be no need for the huge expenditures associated with environmental regulation.

Risk assessment refers to the technical assessment of the nature and magnitude of risk and the goal of risk assessment is to estimate the severity and likelihood of

harm to human health resulting from exposure to a substance or activity that under plausible circumstances can cause harm to human health [2]. Risks are all around us. There are natural forms of risk such as tornadoes, earthquakes, lightning, and disease, or man-made forms such as motorcycles, automobiles, vaccinations, and from production. There are risks that combine natural and man-made causes such as the risk a human takes when driving an automobile from one location to another at night during a rainstorm, the risk that a skyscraper might be damaged by an earthquake, or the risk that we will eat chemicals in our food.

Life is already unacceptably hazardous, and likely to grow more so as the result of excessively rapid and overwhelming technological change while the public consensus, if there is one, seems to be that risk-taking, like religion, travel, or marriage, should be a private affair [3]. Indeed, consumer hostility to public risk is matched only by consumer affection for private risk, in such cases are the aerial sprayings of malathion in a California program to combat the Mediterranean fruit fly provokes passionate opposition, but consumers eagerly spray tens of thousands of gallons of the same pesticide in their own, private gardens; a proposal to ban saccharin precipitates a panic buying of the sweetener; mass airplane accidents cause great concern, and the mandatory use of seat belts in planes is accepted without any problem; but in a much more hazardous private-risk environment of automobile travel, seat belt interlock systems or mandatory seat belt laws have encountered vociferous consumer opposition [4].

The dominant popular attitude seems to reflect a strong aversion to public risk, as opposed to private risk alternatives, in such cases as individuals protesting when required to fasten seat belts in their cars but obeying requirements to buckle up in commercial airplanes; they object to mandatory vaccination but insist on a right of access to incompletely tested miracle cures; they fear additives alleged to be carcinogenic, and yet they consume natural foods that contain carcinogens [5]. States such as Indiana and California do not have a mandatory motorcycle helmet law, even though there is an increased risk of injury if an individual is involved in an accident and is not wearing a helmet. This private risk is accepted by society. Studies indicate that a media campaign encouraging the wearing of motorcycle helmets might save a life for \$3,000 [6].

It seems utterly amazing that we as a society, whether through government regulation or business activity, will spend millions to try to save a life that might be exposed to a public risk, but, in the past, has not been willing to spend the \$3,000 that might save a life exposed to a private risk. Our culture is so closely linked to the concept of personal rights that we contend that individuals have the right to carry on almost any activity that does not injure someone else. In many instances, individuals only look at the direct injury, not the indirect injuries. The fear of risks often

influences perceptions of how great a specific risk really is. Individuals tend to most fear risks that are out of their control and that have severe consequences. Individuals focus on the very small chance of being in an airplane crash or of getting cancer from pesticides. Yet, large risks that individuals can control are often ignored.

A Harvard study indicates that 59% of Americans are worried about eating fruits, vegetables and other foods that are covered with pesticides, and 40% are worried about getting cancer from electrical power lines or cellular phones [7]. For individuals who have such concerns, these are serious risks, but they are not realistic since the number one and three killers are heart disease and stroke, number two is cancer, number four is lung disease, number five is accidents, number six is pneumonia/influenza, number seven is diabetes, number eight is HIV/AIDS, number nine is suicide, and number ten is liver disease [8]. Over a lifetime, the likelihood of dying in a motor vehicle accident is one in 60, which compares to a one in 200,000 chance of dying from pesticide poisoning, a one in 700 chance when living with a smoker, and a one in six chance caused by smoking a pack of cigarettes a day [9].

Of the 2.2 million Americans who die each year, about 22 percent, or 500,000, die of cancer and the number of deaths that are caused by exposure to substances that the government does or might regulate, such as chemical pesticides, various pollutants, or food additives, is the subject of considerable scientific dispute [10]. Two leading authorities, Richard Doll and Richard Peto, published findings about the causes of cancer deaths that suggest that pollution and industrial products account for under three percent, or less than 15,000 deaths, and occupation accounts for four percent, or 20,000 of all cancer deaths [11].

Other related scientific work indicates that substance exposure may account for up to ten percent, or 50,000 death with the range of expert estimates being roughly 10,000 to 50,000 deaths, and experts believe that only a relatively small portion of non-occupational cancers can be regulated at all [12]. By way of comparison, smoking-related cancer accounts for thirty percent, or about 150,000, of those 500,000 deaths, while the number of individuals who die each year from types of cancer whose incidence seems likely to be reduced by regulation is below an estimate that varies from between 10,000 and 50,000, it is probably less than two percent to ten percent of all cancer death and amounts to between seven percent and 33 percent of deaths associated with smoking [13]. Thus, the number of deaths from sources that are regulated must range from less than one percent to less than three percent of the 2.2 million Americans that die each year [14].

An individual might indicate that the private risk of riding a motorcycle without a helmet or not fastening a seat belt does not create a risk or injury to anyone but that individual. Unfortunately, that is not correct. These private risks indirectly impact others in society, whether they do so through the increased cost of insurance because

of these individuals' injuries or through individuals who do not have insurance or assets to pay their medical bills, causing increased costs to be passed indirectly to the rest of society through increased taxes. Other indirect costs could relate to the use of emergency rooms, hospital rooms, and health care providers that are not available immediately to provide medical care to others.

In regard to public risks, the...National Research Council of the National Academy of Sciences and the Environmental Protection Agency outline four steps for most risk assessments. The first step is hazard identification, or initial evaluation of the chemical for health effects. Compounds are particularly studied for genetic mutations, and damage to fetuses. The next step is dose-response assessment, determining potency. The more potent a substance, the less is necessary to produce an adverse health impact. The third step is exposure assessment, evaluating the likelihood of exposure. Exposure may be at a low level, but long-lasting; conversely, it may be a high level, but acute. Different types of exposures may render different effects. In addition, certain individuals may be more prone to exposure. Those closest to a pollution source will probably receive the greatest exposure, for example. Furthermore, exposure to a compound may occur through a number of routes. For example, an air pollutant may be breathed through the lungs, may settle on plants and then ingested, or may settle in water, and be ingested by fish and ultimately consumed by man. Exposure assessment entails characterizing all possibilities of toxicity (step two) with the likelihood of exposure to it (step three) and determines the health risk. This final step accounts for all variables: the toxicity of the compound, the modes of exposure, and variability among individuals who may be exposed [15].

These steps can be used to generate a risk estimate which should be free from biases or values that are typical in private risk assessment. People perceive risks differently depending on (1) the likelihood of a hazard having adverse effects; (2) whom it affects; (3) how widespread, familiar, and dreaded the effects are; (4) how a hazard affects the individuals personally; and (5) whether they have voluntarily agreed to bear the risks [16]. Perceptions of risk are also influenced by the benefits derived from accepting the risks [17]. In considering these perceptions regarding risk, we might also want to look at whether the individual has found out about the risks latent in the individual's conduct, which raises questions about justification or looking solely to the risk, abstracted from the personality of the risk-creator; what are the benefits to society of the risk; what are the costs to society; does the risk maximize utility even if it infringes on an individual's rights [18]? If we deal with justification, the only relevant question is whether the risk on balance is socially desirable [19].

In an economic analysis, rights are seen typically as purely legal entities with instrumental use, rather than as having any intrinsic value [20]. Amartya Sen's book On

Ethics and Economics provides this example:

If person A is violating in a serious way some right of B, e.g. beating him up badly, does person C have a duty to help prevent this? Further, would C be justified in doing some minor violation of some other right of person D to help prevent the more important violation of B's rights by strong-armed A? Could C, for example, take without permission - let us say by force - a car belonging to D who will not lend it to C, to rush to the spot to rescue B from being beaten up by A. If rights only take the form of constraints ('Do not violate the rights of others'), and the constraints are as they are specified in, then C clearly must not try to help B in this way, since C is:

1. under no obligation to help B.
2. he is under an obligation not to violate D's rights [21].

It would be a mistake to ignore consequences and the intrinsic value of any activity is not an adequate reason to ignore its instrumental role, therefore to get an overall assessment of any activity it is necessary to not only look at its intrinsic value, but to also look at its instrumental role and its consequences on other things [22].

2.1 DIFFERING PERSPECTIVES ON PERCEIVED DANGERS: THE GENERAL PUBLIC VS. THE EPA

There is a fundamental difference between the public's perception of risk and the expert's perspective, with that in mind, we can look at how EPA managers provided their own views of program priority rankings and compared them with existing priorities [23]. Subjects the risk managers ranked low, such as hazardous waste cleanup, had high funding priorities; subjects that they ranked high, such as indoor air pollution and global warming, had low funding priorities, and EPA's Science Advisory Board conducted a similar exercise and confirmed the risk manager's views. The public's ranking of safety priorities starts with one as the most important and 26 as having the lowest priority. This is very different from EPA's expert's views, as can be seen by the following:

Public	EPA experts
1. Hazardous waste sites	Medium-to-low
2. Exposure to work site chemicals	High
3. Industrial pollution of waterways	Low
4. Nuclear accident radiation	Not ranked
5. Radioactive waste	Not ranked
6. Chemical leaks from underground storage tanks	Medium-to-low
7. Pesticides	High
8. Pollution from industrial accident	Medium-to-low
9. Water pollution from farm runoff	Medium
10. Tap water contamination	High
11. Industrial air pollution	High
12. Ozone layer destruction	High
13. Coastal water contamination	Low
14. Sewage-plant water pollution	Medium-to-low
15. Vehicle exhaust	High
16. Oil spills	Medium-to-low
17. Acid rain	High
18. Water pollution from urban runoff	Medium

19. Damaged wetlands	Low
20. Genetic alteration	Low
21. Non-hazardous waste sites	Medium-to-low
22. Greenhouse effect	Low
23. Indoor air pollution	High
24. X-ray radiation	Not ranked
25. Indoor radon	High
26. Microwave oven radiation	Not ranked

Regulation generally has been promulgated in response to public outcry over practices that cause harm to the environment, while regulation and the corporate response to regulation have thus been trapped in permanent crisis management, with little private or public trust and a default to antagonism when problems must be solved [25]. Differences between the public's and the expert's evaluation of risks may suggest that the public fears certain risks more than others with the same probability of harm, while in regard to two equal risks, one could rationally dislike or fear more the risk that is involuntarily suffered, public, new, unobservable, uncontrollable, catastrophic, delayed, a threat to future generations, or accompanied by pain and yet, an individual would like to minimize risks of death to himself, family, neighbors, and would normally prefer that regulation buy more safety for a given expenditure or the same amount of safety for less [26]. For example, the public's aversion to toxic water dumps arises because the public does not believe that the risks are low, which means that the public's non-expert reactions reflect not different values, but a different understanding about the underlying risk-related facts and therefore, to change public reaction, one would either have to institute widespread public education in risk analysis or generate greater public trust in some particular group of experts [27].

Consumer Reports indicates that the number of people killed annually in the United States is as follows:

Cause of death	Deaths/year
Tobacco use	*419,000
Alcohol-related	108,000
Motor-vehicle accidents	42,000
Secondhand smoke	*40,000
AIDS	30,000
Drunk driving	17,000
Radon (lung cancer)	*13,600
Food poisoning	*9,000
Falls in the home	7,100
Drowning	4,800
Fires in home	3,200
X-rays	*3,000
Bicycle accidents	950
CO poisoning at home	*200
Hunting accidents	140
Lightning	93
Tornadoes	82
Bees, wasps	64
Skydiving	41
Skiing	41
Football	13
Hang gliding	8

Note: This list omits many other causes of deaths. Those marked * are estimates obtained from reliable sources [28].

Consumer Reports provide information regarding actual

deaths. This information compared with the survey done by the EPA's Science Advisory Board indicates a significant difference between what the public fears are high risk of death exposures and what exposures actually cause death.

2.2 THE TROUBLES DEFINING FAIRNESS AND THE MOVEMENT TO ESTABLISH REASONABLENESS

There are fundamental differences between how lawyers, a person on the street, and an administrator view risk. This raises significant problems and conflicts. Therefore, what about the principle of fairness? Do all individuals in society have the right to roughly the same degree of security from risk? Under our present common law tort system, individuals are subject to harm, without compensation, from background risks. However, no one may suffer harm from additional risks without recourse for damages against the risk-creator, if it can be proven that the individual did not suffer from a background risk or increased risk to a group, but it is more probable than not that a specific risk-creator or group caused injury to that specific individual. Recovery for increased risk of disease has rarely been allowed by the courts, but where it has, recovery of damages has been predicated on proving to a reasonable degree of medical probability, which is greater than fifty percent chance, that the disease will occur [29]. Compensation is a surrogate for the individual's right to the same security as others enjoy, and thus is a violation of an individual's rights to equal security, yet this does not mean that the individual should be able to enjoin the risk-creating activity, since the interests of society may require a disproportionate distribution of risk [30]. The business, scientific, and legal communities are struggling with this situation.

The questions of risk and who subjects whom to excessive risk may come down to reasonableness. Why should the rhetoric of reasonableness and the ability to foresee, appeal to attorneys and judges as a scientific or precise way of thinking, when the paradigm of reasonableness entails several steps of analysis, which involves defining the risks, assessing their consequences, and balancing costs and benefits [31]. The common law tort liability fault requirement shifts the orientation from excusing risks over the rights of individuals to justifying them, which has the following consequences: (1) fault becomes a legal judgment about the risk, rather than a judgment about the responsibility of the individual or firm who created the risk; (2) fault is no longer a question of fairness or rights of the individual, but an inquiry into the relative costs and benefits of the particular risks to society; (3) fault becomes a necessary condition for recognizing the right of the victim to recover [32]. While reasonableness is determined by a simple balancing of costs and benefits, and if the risk yields a net social benefit, traditionally, the

victim's rights are violated and the victim is not entitled to recover from the risk-creator, but if the risk yields a net social cost, the victim is entitled to recover [33]. The reasonableness of the risk determines whether the victim is entitled to compensation and whether the risk-creator ought to be held liable under a common law tort-based fault system.

Learned Hand's test for fault defines the defendant's duty of care as a function of three variables: (1) the probability that the accident will occur, (2) the gravity of the injury which will be suffered if the accident does occur, and (3) the burden of precautions adequate to prevent such accidents. If the cost to the defendant of avoiding the accident would have been less than the cost of the accident, discounted by the probability of its occurrence, the defendant's failure to avoid the accident is termed negligence [34].

3. THE MARKET AND THE ECONOMIST'S PERSPECTIVE ON ACCOUNTABILITY AND REGULATION

Many environmental laws look toward the owner of the property to mend or maintain any environmental damage. In the scope of the business world, environmental damages are examined from an economic viewpoint. In contrast to other disciplines that may view environmental cleanup as a necessity at any cost, businesses would consider a more cost efficient economic approach. Economists and businesses alike utilize marginal analysis and cost benefit analysis to determine the preferred method of environmentally sound operations. A.C. Pigou and Ronald Coase incorporate economic thought when addressing the question of externalities, as it refers to property rights and the economic liability of pollution.

The market is a competition to acquire a scarce amount of resources in the quest to fulfill unlimited human wants and desires, while scarcity in the production of a commodity regarding resources, (land, labor, and capital) was a prevailing factor even before the Industrial Revolution, but once the emphasis was placed upon a competition in mass production between industries, another problem arose [35]. The production of products also entailed the production of smoke and dust, and this cost of production was no longer solely composed of internal expenditures, it now included external costs, such as health risks and lingering black particulate [36]. These costs are considered externalities, since they are not included in the price of the commodity produced [37]. When these activities result, there is uncertainty, in regard to compensation for the victims.

In 1920, Pigou tackled this issue in *The Economics of Welfare* where he described externality as being the main factor in the divergence of private net product and social net product [38]. As in marginal analysis, in order for the private and social values to be maximized, they must be

equal and to accomplish this Pigou devised a scheme of taxes and bounties that would equate the two variables, causing firms associated with externalities to produce a decreased level that would maximize benefits [39].

To illustrate his point, Pigou looked at the problems caused by automobiles. In a pursuit to satisfy their own needs, drivers crowd onto highways causing delays, congestion and inevitably more pollution. This inconvenience, caused by those with profit seeking motives, is therefore a cost to other drivers. In order to alleviate the congestion and increase the flow of traffic, a tax or a toll could be placed on all drivers. This tax would penalize those who were creating the nuisance and who were victimizing others as a result of their behavior. Under the rule of liability, the government forces the risk-creators to pay the damages. According to Pigou, those who viewed their money as an opportunity cost not worth foregoing could then seek other forms of transportation [40]. Not only was Pigou not able to develop a manner in which such taxes and bounties should be calculated for each risk creator, but also the general consensus of a charge was considered a market failure therefore, Pigou's idea was found to be economically inefficient because of its implications [41]. The traditional example is that those who lived near a steel mill were to be monetarily compensated for the pollution the mill caused, many would move to that area and endure the smog simply to reap the rewards of additional income, and there is no incentive for the victims to protect themselves from emissions, perhaps by relocating, even if it would be more economical than forcing the risk-creator to compensate [42].

A more decentralized approach was later developed that did not necessitate a collection of taxes by the government. This theory became a matter of paying for each gain and each loss. The article "The Problem of Social Cost," published in 1960 by Ronald Coase, viewed pollution through the eyes of an economist utilizing the market of supply and demand. According to Coase, in a world of perfect competition, perfect information, and zero transactions costs, the allocation of resources will be efficient and invariant with respect to legal rules of liability [43]. When externalities are involved, it is often difficult to determine who possesses property or individual rights, and these rights are determined by a court of law, Coase suggests that it would be most beneficial for the two parties to negotiate a solution [44]. Under competitive bargaining, the markets force the risk-creators to pay damages, instead of relying on the government as in Pigou's theorem. Consequently, in an effort to maximize their outcome in the bargaining process, it is not uncommon for one party to unfairly underestimate its benefits in order to minimize its costs.

Coase focuses on costs, inserting that when A demands more of a scarce resource by polluting clean air, it imposes a cost on B, and similarly, when B demands clean air, it imposes a cost on A, therefore in this assumption, the costs

and externalities affect the risk-creator and the victim of pollution equally, for example, if the risk-creator dumped hazardous waste from his business into the river upstream, this would impose a cost on the victim who collects his drinking water downstream, and since it would cost the victim to gather water from another source, the victim would thus demand that the risk-creator not dump his pollutants in the water, yet, changing the risk-creator's methods of disposal would also impose a cost upon the risk-creator, therefore, both parties are the cause of the problem and both must contribute to its resolution [45].

This argument exemplifies how those in the field of economics and business may approach questions of fault and responsibility. Other disciplines indicate that the victim should not have to pay to clean or use the river, but Coase's argument suggests that if the victim demands more of the scarce resource (the water) then his price or cost will rise. Economics, as demonstrated in Coase's theory, concentrates on the usefulness of the river instead of the pollution that exists in the river. Traditional economics dealt with the allocation of desired commodities, but now it must determine how to allocate an undesirable good – pollution.

Just as Pigou had his faults, Coase's theorem has not passed without scrutiny. There are several elements that have been omitted in Coase's considerations [46]. The first is that one party may not have the means to contribute enough money to alter the behavior of the other, thus the victim may not be able to compensate the risk-creator enough for him to be able to dump his waste elsewhere, and second, Coase allows for practicality in that he guarantees efficient but not constant results, due to the changing of property rights and allocation of responsibility, but unfortunately, this inconsistency distracts from the powerfulness of the theorem [47]. Moreover, in order for the theorem to remain constant, several assumptions must be made: that zero income elasticity of demand exists, that products and externalities are joint products, and that rents must exceed any required payments [48]. Coase's theorem, although a vast improvement over Pigou's, has been proven fallible.

Coase and Pigou are both concerned with contractual and legal aspects of property rights, and both assume that institutions are able to function with or without transaction costs. However, Pigou argues that the government should be responsible for enforcing the risk-creator to pay for pollution damages, whereas Coase assures that through a process of bargaining, the market will seek justification from both parties who impose costs. Despite their differences, both view the environment as a scarce resource and attempt to allocate responsibility to those who selfishly abuse it. Coase and Pigou were wise to recognize the need to employ economics in an effort to slow a detrimental occurrence in the market that may be an inevitable, environmental shortage.

3.1 COST BENEFIT ANALYSIS AND PARETO

SUPERIOR

Only after risk assessment is conducted can a threshold be determined. Only after the threshold is met do we get to cost-benefit analysis. But what standards should be used for cost-benefit analysis? We could use Pareto superior, Pareto optimal, Pareto superior plus compensation, and/or Executive Order 12291.

These theories can be thought as rights, in regard to the individual pursuing anything the individual wishes, provided the individual does not violate the constraints that restrain individuals from interfering in the legitimate activities of another [49]. In our society, the individual is free to pursue self-interest without hindrance and the existence of such a right restrains others from stopping this individual, if he or she were to pursue self-interest maximization [50].

A Pareto superior outcome is one in which, in comparison to some original position, no one in his own estimation is worse off and at least one person is, in his own estimation, better off [51]. An understanding of economic efficiency begins with Pareto optimality, and in Pareto optimal allocation, there cannot be a reallocation of resources to improve one person's welfare without impairing at least one other person's welfare [52]. In other words, Pareto optimality indicates that there cannot be a change without making at least one person worse off, and Pareto optimal is usually what economists refer to when they deal with the economic efficiencies of our market system, which in the market system involves making exchanges (purchases) under Pareto superior [53]. The consumer continues to purchase items such as bread, milk, or gasoline and the merchant continues to supply them until either the customer or the merchant is no longer better off and at this point, exchange stops because Pareto optimality has been reached [54]. If it costs nothing to get all the individuals in an economy together and to have them reach an agreement about how the economy's resources should be employed, and if all these individuals are rational, then necessarily, they would agree upon a Pareto-efficient allocation [55]. Under the market system, at Pareto optimality, the optimal economic state of affairs has been reached.

The idea that overall social optimality must require Pareto optimality is based on the notion that if a change is advantageous for each, then it must be a good change for the society, yet this economic efficiency pays no attention to distributional considerations involving utility [56]. A state can be Pareto optimal with some people in extreme misery and others rolling in luxury, so long as the miserable cannot be made better off without affecting the luxury of the rich, therefore using Pareto optimality as the only criterion of judgment and self-interest behavior as the only basis for economic choice is a limited way of assessing social achievement [57].

If a society accepts a subjective sense of satisfaction and

consumer sovereignty as its standards, economists can show that society should be interested in achieving economic efficiency [58]. These Pareto theories provide the consumer with the right to choose and have great appeal, since they describe the conditions for free exchange in a market system; therefore, there has been a movement to use the market system theory as an economically efficient guide for government regulation and interference with free exchange.

Under the market system approach, economists assume that the potential customer has perfect information, but in relation to toxins such as benzene, radon, diethylstilbetrol, formaldehyde, and others, potential customers do not have perfect information, and actual consent may not be meaningful. This complication has caused the government to move from actual consent, the right to choose under the market system in which the consumer makes all the choices on consumption, to the government's assumption of what choices the citizen would make if the consumer had perfect information. This movement causes difficulties because of the 313 million citizens of the United States. Given self-interest behavior, the market mechanism provides good incentives for consumers to choose appropriately, given their individual endowments, but there is no comparable mechanism by which individuals have the incentive to reveal information outside of the market system, and any departure from self-interest behavior may threaten the achievement of Pareto optimality [59]. Without the availability of Pareto optimality, costs can be estimated fairly well, but benefits cannot be because they are very subjective. Thus, we infer other people's benefits—often based on our own. Since the government's contention is that actual consent may not be meaningful, it has not been inclined to survey a representative sample of citizens in relation to increased risk. It does use such a survey for unemployment figures by sampling 50,000 individuals and indicating with significant confidence the unemployment rate. In order to be economically efficient, we must find Pareto optimality, which is our starting point, and this goal is difficult to achieve outside of the market system and without reducing the sample size from 313 million.

3.2 COST-BENEFIT ANALYSIS AND PARETO SUPERIOR PLUS COMPENSATION

If government is striving to be economically efficient, and the assumption is that the market system is efficient, then either the classical definitions of Pareto superior and Pareto optimality may need to be changed or the application needs to be changed. Such change would come as another recognition that in an imperfect world, the right to choose may be eliminated and thus Pareto optimality can be approached more closely by systems which involve some coercion, rather than by a system of totally free bargains [60]. Optimality may be difficult to achieve in a world with high transaction costs [61]. Most voluntary

transactions under property rights move property from lower to higher value uses, which means that the purchase price is greater than the value of the property to the seller, or else the seller will not part with it, yet the purchase price must be less than the value of the property in use to the buyer, or else the buyer will not purchase it [62]. When total transaction costs exceed the difference in values to the buyer and the seller, then the exchange cannot go forward, since both parties no longer will emerge as net winners, since success in encouraging voluntary transaction therefore lies in the reduction of transaction costs [63]. Since such costs cannot always be reduced, we need to give the person something to compensate him or her after the fact. This compensation would be similar to a worker's rights under Workers' Compensation in which government tries to estimate benefits. If one loses an arm, leg or life, either his or her estate is compensated for it after the fact. The huge assumption is that no one is made worse off because of the compensation after the fact. This system is a good example of Pareto superior plus compensation.

Under this theory, how economically efficient is it if each individual who is affected by a hazardous or toxic substance receives compensation? As an example, I will list a few of the substances and/or situations I come in contact with in an average day: I smell benzene when putting gasoline in my car; I smell exhaust fumes from a city bus as I am about to cross the street; there is radon in my home; there are pesticides on my food; there is freon in my air conditioner; there is lead solder from the pipes in the water that sits overnight; there is a glue smell from the linoleum floor; there is Nutra-Sweet in my Diet Coke; a commercial dumpster I walk by gives off strong odors; etc.

Pareto superior plus compensation would involve, in essence, tort reform, and the government has not moved beyond common law tort liability for environmental increased risk to strict liability. Therefore, if the market-type system is going to be used, government has determined that the actual compensation would be watered down to potential compensation. Potential compensation is not a Pareto superior move. There are no longer winners and winners, as in the market system in which everyone benefits from the exchange. There are winners who could potentially compensate the losers, but winners still have something left over. This standard of economic efficiency could be used to determine what creates the most wealth in situations in which the winners could potentially compensate the losers. This standard, then, ties right into cost-benefit analysis. Historically, cost-benefit analysis has been associated with government public works construction projects in which the ratio of benefits to costs must exceed 1 to 1 in order for a project to be funded. The idea behind cost-benefit analysis is to compare alternative courses of action, which include the option of doing nothing and continuing the status quo.

The United States Supreme Court, in *American Textile Manufacturers Institute, Inc. v. Donovan*, dealt with the

issue of cost-benefit analysis in regard to toxic substances [64]. The court determined, in regard to occupational safety and health standards, that feasibility would involve economical feasibility, not just technical feasibility, and that cost is a constraint. The Court did not issue a Delaney Clause standard or a threshold standard, but rather a cost-benefit analysis standard.

Closely related to the Supreme Court's decision was Executive Order 12291 issued in 1981 by President Reagan. This Executive Order was a mandate to all agencies, to the extent permitted by law, that...regulatory action shall not be undertaken unless the potential benefits to society for the regulation outweigh the potential costs to society; regulatory objectives shall be chosen to maximize the net benefits to society; among alternative approaches to any given regulatory objective, the alternative involving the least net cost to society shall be chosen; and agencies shall set regulatory priorities with the aim of maximizing the aggregate net benefits to society, taking into account the condition of the particular industries affected by regulations, the condition of the national economy, and other regulatory actions contemplated for the future [65].

Note that the Executive Order states to the extent permitted by law, and this qualification is included because Congress may forbid an agency from considering costs (even though benefits may not outweigh costs in all cases) because the transaction costs of performing a formal analysis exceed the benefits, and Congress may perform an intuitive cost-benefit analysis, estimating that the benefits of minimum health standards outweigh the costs of attaining them, or that agency consideration of benefits and costs is preempted by Congress's political assessment of the balance [66].

Cost-benefit analysis is a frequently employed input to decisions about setting emission limits since the 1980s. Regulatory reformers, however consider cost-benefit analysis a form of priority setting, because they imagine that stringency determinations divert resources from some health and safety priorities to others [67]. But they do not require agencies to weigh those costs against the value of avoided harms (usually referred to as benefit) [68]. In *Chemical Manufacturing Association v. EPA*, the court concluded that the Administrator must determine whether costs can be reasonable borne by industry, but need not conduct a cost-benefit analysis [69].

4. CONCLUSION

So far in this article, we have touched upon many questions regarding standards for environmental regulations: What should drive environmental law decisions? Which environmental concerns present the greatest risk? Which risks should be targeted? Should our judicial system even have the power to regulate these risks? Should proof of causation be required in order to establish regulations, and if so, how do we determine causation? How do we place value on life?

The partnership of scientific research and the legal system is becoming increasingly more important with the rising complexity of environmental issues. Current environmental issues on trial require a far deeper erudition about chemical reactions, the function of ecosystems, and the limits to safe exposure—erudition which outreaches the scope of knowledge possessed by traditional lawyers. Science is concerned with objective truths and theories that can be verified by those who have extensive training in scientific method. On the other hand, law involves rules and regulations translated by judges and juries, who lack the knowledge to adequately and thoroughly evaluate scientific evidence.

REFERENCES

- [1] J. Kim, "Preemption After Cipollone: Preserving State Tort Remedies in the Face of Federal Regulation," *University of Toledo Law Review*, vol. 20, pp. 797, 845, 1989.
- [2] J. Cohns and V. Covello, *Risk Analysis: A Guide to Principles and Methods for Analyzing Health and Environmental Risks*, Executive Office of the President of the United States, Council on Environmental Quality, pp. 1-3, 55, 1989.
- [3] Z. Plater, R. Abrams and W. Goldfarb, *Environmental Law and Policy: Nature, Law, and Society*, pp. 82-83, West, 1992.
- [4] Plater, Id. p. 83.
- [5] C. Gillette and J. Krier, "Risk, Courts, and Agents", 138 *University of Pennsylvania Law Review*, 1027, 1032, 1990.
- [6] S. Rhoads, "How Much Should We Spend to Save a Life?", 51 *The Public Interest*, pp. 74, 79, 1978.
- [7] M. Ryan, "What's Really Risky," *Parade Magazine*, 16, June 15, 1997.
- [8] M. Ryan, Id. p. 16.
- [9] J. Walsh, *True Odds—How Risk Affects Your Everyday Life*, 16, Merritt Publishing, 1996.
- [10] S. Breyer, *Breaking the Vicious Circle: Toward Effective Risk Regulation*, Harvard University Press, 6, 1994.
- [11] S. Breyer, Id. p. 6.
- [12] S. Breyer, Id. p. 6.
- [13] S. Breyer, Id. pp. 6, 7, and 8.
- [14] S. Breyer, Id. p. 8.
- [15] H. Doliner, "The Use of Risk Assessment in Resolving Environmental Controversies," 4 *Toxics Law Reporter*, 717, Nov. 15, 1989.
- [16] J. Cohns and V. Covello, "Risk Analysis: A Guide to Principles and Methods for Analyzing Health and Environmental Risk," Executive Office of the President of the United States, Council on Environmental Quality, p. 9, 1989.
- [17] Cohns, Id. p. 9.
- [18] G. Fletcher, "Fairness and Utility in Tort Theory," 85 *Harvard Law Review*, 537, 559, 1972.
- [19] G. Fletcher, Id. at 559.
- [20] A. Sen, *On Ethics and Economics*, p. 71. Basil Blackwell, 1990.
- [21] A. Sen, Id. p. 71-73.
- [22] A. Sen, Id. p. 75.
- [23] S. Breyer, supra note 10 p. 20.
- [24] S. Breyer, supra note 10, pp. 20-21.
- [25] M. Epstein, *Measuring Corporate Environmental Performance*. Irwin, X, 1996.
- [26] S. Breyer, supra note 10, pp. 33 & 35.
- [27] S. Breyer, supra note 10, pp. 33 & 39.
- [28] "Facing Our Fears," 61, *Consumer Reports*, 52, December 1996.
- [29] J. Weinstein and E. Hershenov, "The Effect of Equity on Mass Tort Law," *University of Illinois Law Review*, vol. 2, 269, 313, 1991.
- [30] G. Fletcher, supra note 18, p. 550.
- [31] G. Fletcher, supra note 18, pp. 570 and 573.
- [32] G. Fletcher, supra note 18, pp. 559 & 560.
- [33] G. Fletcher, supra note 18, p. 542.
- [34] G. Calabresi and J. Hirschoff, "Toward a Test for Strict Liability in Torts," *Yale Law Journal*, vol. 81, no. 6, pp. 1055-1085, 1972.
- [35] Laftont, "Externalities" in *The New Palgrave Dictionary of Economics*. Macmillan, 263, 1987.
- [36] Laftont, Id. p. 263.
- [37] Laftont, Id. p. 263.
- [38] S. Lin, *Theory and Measurement of Economic Externalities*. Academic Press, 2, 1976.
- [39] S. Lin, Id. p. 2.
- [40] A. C. Pigou, *The Economics of Welfare*. Macmillan, pp. 192-193, 196.
- [41] W. Baumol and W. Oates, *The Theory of Environmental Policy*. Prentice-Hall, p. 25, 1975.
- [42] W. Baumol and W. Oates, Id. 25-26.
- [43] S. Lin, supra note 38, p. 29.
- [44] S. Lin, supra note 38, p. 3.
- [45] W. Baumol and W. Oates, supra note 41, at 31.
- [46] S. Lin, supra note 38, p. 38.
- [47] S. Lin, supra note 38, p. 38.
- [48] S. Lin, supra note 38, p. 38.
- [49] A. Sen, supra note 20, p. 56.
- [50] A. Sen, supra note 20, p. 56.
- [51] T.S. Ulen, "The Efficiency of Specific Performance: Toward a Unified Theory of Contract Remedies", 83 *Michigan Law Review*, vol. 83, no. 2, pp. 341-403, 1984. See H. Kohler, *Intermediate Microeconomics*. Scott Foresman, pp. 417-436, 1982.
- [52] S.E. Rhoads, *The Economist's View of the World: Government, Markets, & Public Policy*. Cambridge University Press, p. 63, 1985.
- [53] S. Rhoads, Id. p. 63.
- [54] S. Rhoads, Id. p. 63.

- [55] D.H. Regan, "The Problem of Social Cost Revisited," *J. Law & Economics*, vol. 15, no. 2, pp. 427-437, 427, 428-429, 1972.
- [56] A. Sen, supra note 40, pp. 33 & 38.
- [57] A. Sen, supra note 40, pp. 32-35.
- [58] S. Rhoads, supra note 122, p. 63.
- [59] A. Sen, supra note 40, pp. 36-37 and 53.
- [60] G. Calabresi and A.D. Melamed, "Property Rules, Liability Rules, and Inalienability: One View of the Cathedral," *Harvard Law Review*, vol. 85, no. 6, pp. 1089-1128, p. 1113, April 1972.
- [61] S. Rose-Ackerman, "Inalienability and the Theory of Property Rights", *Columbia Law Review*, vol. 85, no. 5, pp. 931-969, 1985.
- [62] R.A. Epstein, "Why Restrain Alienation," *Columbia Law Review*, vol. 85, no. 5, pp. 970-990, 1985.
- [63] R.A. Epstein, Id, p. 972.
- [64] American Textile Manufacturers Institute, Inc. v. Donovan, 452 U.S. 490, 1981.
- [65] Executive Order 12291, 46 FR 13193, 17 Feb. 1981.
- [66] J.E. Bonine and T.O. McGarity, *The Law of Environmental Protection: Cases*, Westlaw, 2nd Ed., p. 668, 1992.
- [67] D.M. Driesen, "Distributing the Costs of Environmental Health, and Safety Protection: The Feasibility Principle, Cost-Benefit Analysis, and Regulatory Reform," *Boston College Environmental Affairs Law Review*, vol. 32, no. 1, pp. 1-95, 2005.
- [68] EPA v. National Crushed Stone Association, 449 U.S. 64, 71 1980.
- [69] Chemical Manufacturers Association v. EPA, 870 F2d 177, 262, 1989.