

INTERNATIONAL JOURNAL OF SCIENTIFIC AND ENGINEERING RESEARCH

IJSER

VOLUME 2, ISSUE 4, APRIL 2011

ISSN 2229-5518




**Research
Publication**

IJSER

<http://www.ijser.org>

<http://www.ijser.org/xplore.html>

<http://www.ijser.org/forum>

E-mail: ijser.editor@ijser.org

ISSN 2229-5518



9 772229 551823



04

International Journal of Scientific and Engineering Research (IJSER)

Journal Information

SUBSCRIPTIONS

The International Journal of Scientific and Engineering Research (Online at www.ijser.org) is published monthly by IJSER Publishing, Inc., France/USA/India

Subscription rates:

Print: \$50 per issue.

To subscribe, please contact Journals Subscriptions Department, E-mail: sub@ijser.org

SERVICES

Advertisements

Advertisement Sales Department, E-mail: service@ijser.org

Reprints (minimum quantity 100 copies)

Reprints Co-ordinator, IJSER Publishing.

E-mail: sub@ijser.org

COPYRIGHT

Copyright©2011 IJSER Publishing, Inc.

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as described below, without the permission in writing of the Publisher.

Copying of articles is not permitted except for personal and internal use, to the extent permitted by national copyright law, or under the terms of a license issued by the national Reproduction Rights Organization.

Requests for permission for other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works or for resale, and other enquiries should be addressed to the Publisher.

Statements and opinions expressed in the articles and communications are those of the individual contributors and not the statements and opinion of IJSER Publishing, Inc. We assume no responsibility or liability for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained herein. We expressly disclaim any implied warranties of merchantability or fitness for a particular purpose. If expert assistance is required, the services of a competent professional person should be sought.

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:

E-mail: ijser.secretary@ijser.org

Contents

1. **Haptic Training Method Using a Nonlinear Joint Control**
Jarillo-Silva Alejandro, Domínguez-Ramírez Omar A., Parra-Vega Vicente.....1-8
2. **Prevalence of Eating Disorder among Female Students of Tonekaban University**
Ida Fadavi, Roxanna Ast.....5-13
3. **Impact Of Leverage On Firms Investment Decision**
Franklin John. S, Muthusamy13-28
4. **Effect of Nanofluid Concentration on the Performance of Circular Heat Pipe**
M. G. Mousa29-36
5. **Enhancement of person identification using Iris pattern**
Vanaja roselin.E.Chirchi, Dr.L.M.Waghmare, E.R.Chirchi.....37-42
6. **Mn DOPED SnO₂ Semiconducting Magnetic Thin Films Prepared by Spray Pyrolysis Method**
K.Vadivel, V.Arivazhagan, S.Rajesh43-47
7. **A novel comprehensive method for real time Video Motion Detection Surveillance**
Sumita Mishra, Prabhat Mishra, Naresh K Chaudhary, Pallavi Asthana.....48-51
8. **Handoff Analysis for UMTS Environment**
Pankaj Rakheja, Dilpreet Kaur, Amanpreet Kaur.....52-55
9. **Performance and Emission Characteristics of Stationary CI Engine with Cardnol Bio fuel blends**
Mallikappa, Rana Pratap Reddy, Ch.S.N.Muthy56-61
10. **Morphological Space and Transform System**
Ramkumar P.B, Pramod K.V.....62-66
11. **Bus Proximity Indicator - (An Intelligent Bus Stop)**
Prof. A.P. Thakare, Mr. Vinod H. Yadav67-70
12. **Designing aspects of Artificial Neural Network controller**
Navita Sajwan, Kumar Rajesh71-75

Contents

13. Handwritten Character Recognition using Neural Network Chirag I Patel, Ripal Patel, Palak Patel	76-81
14. Improved Performance of M-ary PPM in Different Free-Space Optical Channels due to Reed Solomon Code Using APD Nazmi A. Mohammed, Mohammed R. Abaza and Moustafa H. Aly	82-85
15. Electricity Sector Restructuring Experience of different Countries Archana Singh, Prof. D.S.Chauhan	86-93
16. Significant Role of Search Engine in Higher Education Rahul J. Jadhav, Dr. Om Prakash Gupta, Usharani T. Pawar	94-98
17. Feature Selection for Cancer Classification: A Signal-to-noise Ratio Approach Debahuti Mishra, Barnali Sahu.....	99-105
18. Towards Analysis of Vulnerabilities and Measuring Security Level of Copmputer Network Life Cycle Irshad Ahmad Mir, Mehraj-U-Din Dar, S.M.K Quadri.....	106-111
19. Text Independent Speaker Identification In a Distant Talking Multi-Microphone Environment Using Generalized Gaussian Mixture Model P. Soundarya Mala, Dr. V. Sailaja, Shuaib Akram.....	112-116
20. Novel Object Removal in Video using Patch Sparsity B. Vidhya, S. Valarmathy.....	117-121
21. New formula of Nuclear Force Md. Kamal uddin.....	122-128
22. Impact Fatigue Behaviour of Fully Dense Alumina Ceramics with Different Grain Sizes Manoj Kumar Barai, Jagabandhu Shit, Abhijit Chanda, Manoj Kr Mitra.....	129-133
23. Distributed Generation Planning Optimization Using Multiobjective Evolutionary Algorithms Mahmood Sheidaee, Mohsen Kalantar.....	134-139

Contents

- 24. Marketing Of Asian Countries as Tourist Destination- Comparative Study of India and Malaysia**
Dr Reshma Nasreen, Nguyen Toan Thang.....140-156
- 25. Clinical and Computational Study of Geometry & Heamodynamics of Arterial Stenosis**
Krittika Dasgupta, Abhirup Roy Choudhury, Abhijit Chanda, Debabrata Nag.....157-162
- 26. Segmentation Techniques for Iris Recognition System**
Surjeet Singh, Kulbir Singh.....163-170
- 27. Investigation of X-ray Plasmon Satellites in Rare Earth Compounds**
Ajay Vikram Singh, Dr. Sameer Sinha.....171-175
- 28. Improvement of Power System Stability by Simultaneous AC-DC Power Transmission**
T.Vijay Muni, T.Vinoditha, D.Kumar Swamy.....176-181
- 29. A Comprehensive Distributed Generation Planning Optimization with Load Models**
Mahmood Sheidaee, Mohsen Kalanter.....182-187
- 30. Positive Impact of Mobility Speed on Performance of AODV at Increased Network Load**
Rajesh Deshmukh, Asha Ambhaikar.....188-192
- 31. Development and Physical, Chemical and Mechanical Characterization of Doped Hydroxyapatite**
Promita Bhattacharjee, Howa Begam, Abhijit Chanda.....193-200
- 32. A chart for general and complicated data visualization**
M Siluvairajah.....201-202
- 33. Application of Reliability Analysis: A Technical Survey**
Dr. Anju Khandelwal.....203-210
- 34. An Innovative Quality Of Service (Qos) Based Service Selection For Service Orchestration In Soa**
K.Vivekanandan, S.Neelavathi.....211-218

Contents

- 35. Design of Boost Circuit for Wind Generator**
N.Prasanna Raj, M.Mohanraj, Rani Thottungal.....219-224
- 36. A Novel Method for Fingerprint Core Point Detection**
Navrit Kaur Johal, Amit Kamra.....225-230
- 37. Robust Impulse Eliminating Internal Model Control of Singular Systems**
M.M. Share Pasand, H.D. Taghirad.....231-238
- 38. An Integrated Partial Backlogging Inventory Model having weibull demand and variable deterioration rate with the effect of trade credit**
P.K. Tripathy, S. Pradhan.....239-243
- 39. Survey of Current and Future Trends in Security in Wireless Networks**
Vikas Solomon Abel.....244-249
- 40. Anomalous hydrogen production during photolysis of NaHCO₃ mixed water**
Muhammad Shahid, Noraih Bidin, Yacoob Mat Daud, Muhammad Talha.....250-254
- 41. Peripheral Interface Controller based the Display Unit of Remote Display System**
May Thwe Oo.....255-263
- 42. Training and Analysis of a Neural Network Model Algorithm**
Prof Gouri Patil.....264-268
- 43. Wireless Sensor Network: A Review on Data Aggregation**
Kiran Maraiya, Kamal Kant, Nitin Gupta.....269-274
- 44. Slant Transformation As A Tool For Pre-processing In Image Processing**
Nagaraj B Patil, V M Viswanatha, Dr. Sanjay Pande MB.....275-281
- 45. Separation of concerns in VoiceXML applications**
Sukhada P. Bhingarkar.....282-285
- 46. Error of Approximation in Case of Definite Integrals**
Rajesh Kumar Sinha, Satya Narayan Mahto, Dhananjay Sharan.....286-289

Contents

- 47. An Assessment model for Intelligence Competencies of Accounting Information Systems**
Mehdi Ghazanfari, Mostafa Jafari, Saeed Rouhani.....290-296
- 48. Processing of Images based on Segmentation Models for Extracting Textured component**
V M Viswanatha, Nagaraj B Patil, Dr. Sanjay Pande MB.....297-303
- 49. An Adaptive and Efficient XML Parser Tool for Domain Specific Languages**
W. Jai Singh, S. Nithya Bala.....304-307

Haptic Training Method Using a Nonlinear Joint Control

Jarillo-Silva Alejandro, Domínguez-Ramírez Omar A., Parra-Vega Vicente

Abstract— There are many research works on robotic devices to assist in movement training following neurologic injuries such as stroke with effects on upper limbs. Conventional neurorehabilitation appears to have little impact on spontaneous biological recovery; to this end robotic neurorehabilitation has the potential for a greater impact. Clinical evidence regarding the relative effectiveness of different types of robotic therapy controllers is limited, but there is initial evidence that some control strategies are more effective than others. This paper consider the contribution on a haptic training method based on kinesthetic guidance scheme with a non linear control law (proxybased second order sliding mode control) with the human in the loop, and with purpose to guide a human user's movement to move a tool (pen in this case) along a predetermined smooth trajectory with finite time tracking, the task is a real maze. The path planning can compensate for the inertial dynamics of changes in direction, minimizing the consumed energy and increasing the manipulability of the haptic device with the human in the loop. The Phantom haptic device is used as experimental platform, and the experimental results demonstrate.

Index Terms—Diagnosis and rehabilitation, haptic guidance, sliding mode control, path planning, haptic interface, passivity and control design.

1 INTRODUCTION

IN the least decade the number of patients who have suffered accidents stroke or traumatic brain injuries have increase considerably [1]. The central nervous system damage can lead to impaired movement control upper extremities, which are facing major difficulties in relation to the activities of daily life. Several studies showed that the rehabilitation therapy which is based on motion-oriented tasks repetitive, it help to improve the movement disorder of these patients [2], [3]. Unfortunately repeatability therapy requires consistency, time for physicians and therefore money.

Conventional neurorehabilitation appears to have little impact on impairment on spontaneous biological recovery. Robotic neurorehabilitation has potential for a greater impact on impairment due to an easy deployment, its applicability across of a wide range of motor impairment, its high measurement reliability, the capacity to deliver high dosage of training protocols and high intensity in the exercises. This situation economic rehabilitation. With the purpose of enhance the relationship between outcome and the costs of rehabilitation robotic devices are that being introduced in clinical rehabilitation [4], [5]. Rehabilitation using robotic devices not only has had an important contribution in this area but also has introduced greater accuracy and repeatability of rehabilitation exercises. Accurate measurement quantitative parameter us-

ing robotic instrumentation is a tool that accomplishes the goal of the patient's recovery.

1.1 Justification

As care is decentralized and moves away the inpatient settings to homes, the availability of technologies can provide effective treatment outside, the acute care of the hospital would be critical to achieve sustainable management of such diseases. In the field of neuromotor rehabilitation, skilled clinical managers and therapists can achieve remarkable results without using technological tools or with rudimentary equipment, but such precious human capital is in very short supply and, in an case, is totally insufficient to sustain the current demographic.

Robots are particularly suitable for both rigorous testing and application of motor learning principles to neurorehabilitation. Computational motor control and learning principles derived from studies in healthy subjects are introduced in the context of robotic neurorehabilitation. There is an increasing interest in using robotic devices to provide rehabilitation therapy following neurologic injuries such as stroke and spinal cord injury. The general paradigm to be explored is to use a robotic device to physically interact with the participant's limbs during movement training, although there is also work that uses robot that do not physically contact the patients [6].

1.2 The Problem

Some of the problems presented by a motor rehabilitation of patients who have suffered a brain injury, is the time rehabilitation takes place, the cost that this entails, poor information that the specialist has to determine the diagnosis of a patient in rehabilitation, and the lack of platforms for rehabilitating patients who cannot assist to the hospital

- Jarillo-Silva Alejandro, University of the Sierra Sur, México, Oaxaca, E-mail: ajarilloe@unsis.edu.mx
- Domínguez-Ramírez Omar A. University of Hidalgo State, México, E-mail: omar@uaeh.edu.mx
- Parra-Vega Vicente, Robotics and Advanced, Manufacturing Division, CINVESTAV Saltillo, México, E-mail: vpar-ra@cinvestav.edu.mx

and require continued rehabilitation.

1.3 Our Proposal

One technique used to solve the problem of generating rehabilitation platforms using robotic systems is haptic guidance, this technique is based on the use of haptic devices, which allow human-machine interaction. These haptic devices are programmed under certain considerations, such as considering the safety of the patient during the interaction, the physiology of the patient in order to generate tracking trajectories that allow proper rehabilitation, and so on. In this paper is proposed to generate a trajectory based on the solution of mazes in 2D and the haptic device guides the patient under a control law, which is designed with certain characteristics that allow coupling among the patient, the device and the trajectory. The haptic device is equipped with optical encoders, this allows obtaining data such as position, velocity, acceleration, force and torque, these data in combination with other data allow to the specialist to generate a clinical diagnosis, which has therapeutic support based on the patient's movements in each exercise performed on the platform.

1.3 Organization

In section 2, we introduce the human-haptic interaction, including the haptic scheme, the dynamic model of the haptic device and the guidance control law implemented in the experimental platform. The description task and the path planning for guiding to the experiments are given in section 3; experimental results are discussed in 4. Finally, we present the conclusions in section 5.

3 HUMAN-HAPTIC INTERACTION

The haptic feedback has also been developed for task oriented biofeedback studies [7]. Haptic interfaces allow the patient to interact with and to manipulate a virtual object or real. Results [8] have shown that haptic information provides knowledge of results (KR) and kinaesthetic feedback, sensation that are important for task performance.

3.1 Haptic Guidance Scheme

For the human participation in the system, there are two haptic guidance applications: active and passive. The first allows the user to guide the robotic system (haptic device) in the workspace to explore. However, in the passive haptic guidance, the user is guided in the workspace with the position and velocity control strategies. On the other hand, kinesthesia is the human sense of position and movement, which is created from proprioceptive cues arising from receptors in the joints and muscles [9]. Thanks to kinesthetic memory, or the ability to remember limb position and velocity, humans have a remarkable ability to remember positions of their limb quite accurately and for long periods [10].

3.2 Haptic Device Dynamics

Consider a mechanism of articulate links (PHANTOM 1.0 [11]), with n revolute joints described in the generalized joint coordinates $(q^T, \dot{q}^T)^T \in R^{2n}$. Physically, the robot is never in touch with a physical object, thus robot dynamics in free motion describe properly the haptic device, as follows

$$D(q)\ddot{q} + (C(q, \dot{q}) + B)\dot{q} + G(q) = \tau \quad (1)$$

Where $D(q) \in R^{3 \times 3}$ denotes a symmetric positive definite inertial matrix, $C(q, \dot{q}) \in R^{3 \times 3}$ is a Coriolis and centripetal forces matrix, $G(q) \in R^{3 \times 1}$ models the gravity forces vector, $B \in R^{3 \times 3}$ denotes a symmetric positive definite viscous coefficient matrix, and $\tau \in R^{3 \times 1}$ stands for the torque input. In our case the human operator is driving the system, therefore $\tau = \tau_c + \tau_h$, and $\tau_c \in R^{3 \times 1}$ stands for the guidance control, and $\tau_h = J^T f_h$ denotes the human-haptic interaction, where J^T stands for the transpose Jacobian of the haptic device, and f_h represents the human performance force vector.

3.3 Guidance Control Law

The design of a control strategy applied to haptic guided must meet certain objectives such as: giving stability in human-machine interaction, presenting high robustness against disturbances generated by the human operator and currently meeting the objective to rehabilitate, this can be corroborated with an analysis clinical examined by specialists.

3.4 Properties of Euler-Lagrange Systems

The dynamic equation for manipulator robots (1) has interesting follow properties [12]:

- There exists some positive constant α such that

$$D(q) \geq \alpha I \quad \forall q \in R^n$$

Where I denotes the $n \times n$ identity matrix. The $D(q)^{-1}$ exist and this is positive definite.

- The matrix $c(q, \dot{q})$ have a relationship with the inertial matrix as:

$$\dot{q}^T \left[\frac{1}{2} \dot{D}(q) - C(q, \dot{q}) \right] \dot{q} = 0 \quad \forall q, \dot{q} \in R^n$$

- The Euler-Lagrange system have the total energy as:

$$\mathcal{E} = k + u \quad (2)$$

Where k is the kinetic sum for each coordinate, and u are the potential energy respectively.

The kinetic energy is obtained by:

$$k = \frac{1}{2} \sum_{i=1}^n m_i v_i^2 = \frac{1}{2} \dot{q}^T D(q) \dot{q} \quad (3)$$

Where $m_i \in \mathbf{R}$ are the i mass of the i link, $v_i \in \mathbf{R}^n$ are the i velocity of the i link.

The potential energy is obtained by:

$$u = \sum_{i=1}^n m_i h_i g \quad (4)$$

Where $h_i \in \mathbf{R}^n$ are the i height of the i link respect to mass center and g is a gravitational constant. By differentiating (2) we obtain

$$\begin{aligned} \dot{\varepsilon} &= \dot{q}^T D(q) \dot{q} + \frac{1}{2} \dot{q}^T \dot{D}(q) \dot{q} + \dot{q}^T G(q) \\ &= \dot{q}^T [-C(q, \dot{q}) - G(q) + \tau] + \frac{1}{2} \dot{q}^T \dot{D}(q) \dot{q} + \dot{q}^T G(q) \\ &= \dot{q}^T \tau \end{aligned} \quad (5)$$

- From the passivity property we have that:

$$V(x) - V(0) \leq \int_0^t y^T(s) u(s) ds \quad (6)$$

Where $V(x)$ is a storage function, $y(s)$ is the output, and $u(s)$ is the input of the system, and s is a variable change. Using (5) for the Euler-Lagrange system, energy function ε as the storage function, and we have the passivity property as:

$$\varepsilon(t) - \varepsilon(0) \leq \int_0^t \dot{q}^T \tau dt \quad (7)$$

Where \dot{q} is the output, and τ is the input of the system.

3.5 Passivity of the Dynamic Error

The described passivity property presents an interesting property in the regulation problem on Euler-Lagrange systems, however this property can be extended for the trajectory tracking problem solution, [13] drive this passive error dynamic property, as follow for:

$$D(q) \dot{s} + [C(q, \dot{q}) + K_d(q, \dot{q})] s = 0 \quad (8)$$

Where s denotes an error signal that we want to drive to zero, $k_d(q, \dot{q}) = k_d(q, \dot{q})^T > 0$ is a damping injection matrix. Based on this property and the skew-symmetric property we follow the next lemma:

Lemma 1: The differential equation:

$$D(q) \dot{s} + [C(q, \dot{q}) + K_d(q, \dot{q})] s = \psi \quad (9)$$

Where $D(q)$ and K_d are positive matrices and $C(q, \dot{q})$ satisfies (8) it defines an output strictly passive operator $\sum_d: \psi \rightarrow s$. Consequently, if $\psi \equiv 0$ we have $s \in L_2$

With the Lemma 1, and the dynamic properties, we can follow a control law design. In this paper we obtain the control technique via Lyapunov theory, dynamical and passivity properties.

4 DESIGN OF A NONLINEAR JOINT CONTROL BASED ON PASSIVIT

Given a designated trajectory $q_d(t)$, can be considered as a Lyapunov function as follows [14]:

$$V(s) = \frac{1}{2} s^T D(q) s + \int s^T K_L \int \tanh(s) ds ds + c \quad (10)$$

Where $c \gg 1$, $s = [s_1, s_2, \dots, s_n]^T$ denote dynamic error, the function $\tanh(s) = [\tanh(s_1), \dots, \tanh(s_n)]^T$ and K_L are a symeyric positive definite matrices.

Given that (10) is a positive semidefinite function, obtain the derivative of the Lyapunov candidate function, so that is follows [1]:

$$\begin{aligned} \dot{V}(s) &= \frac{1}{2} s^T D(q) \dot{s} + \frac{1}{2} s^T \dot{D}(q) s + \frac{1}{2} s^T D(q) s \\ &+ s^T K_L \int \tanh(s) ds \end{aligned} \quad (11)$$

Simplifing (11)

$$\dot{V}(s) = s^T D(q) \dot{s} + \frac{1}{2} s^T \dot{D}(q) s + s^T K_L \int \tanh(s) ds \quad (12)$$

As saw in (12), we applied lemma 1, and use the follow result:

$$D(q) \dot{s} = \psi - [C(q, \dot{q}) + K_d] s \quad (13)$$

Where ψ are the system dynamics

$$\begin{aligned} \psi &= \tau - [D(q) \ddot{q}_r + C(q, \dot{q}) \dot{q}_r + G(q) + F(\dot{q})] \\ &= \tau - Y_r(q, \dot{q}, \ddot{q}_r, \dot{q}_r) \phi \end{aligned} \quad (14)$$

Where \dot{q}_r and \ddot{q}_r are the nominal references, $Y_r(q, \dot{q}, \ddot{q}_r, \dot{q}_r) \in \mathbf{R}^{n \times p}$ is the regressor of known nonlinear terms $\phi \in \mathbf{R}^p$ is the vector of unknown parameters, so that (13) is as follows:

$$D(q) \dot{s} = \tau - Y_r(q, \dot{q}, \ddot{q}_r, \dot{q}_r) \phi - [C(q, \dot{q}) + K_d] s \quad (15)$$

Using (15) and (12) as follows:

$$\begin{aligned} \dot{V}(s) &= s^T [\tau - Y_r(q, \dot{q}, \ddot{q}_r, \dot{q}_r) \phi - [C(q, \dot{q}) + K_d] s] \\ &+ \frac{1}{2} s^T \dot{D}(q) s + s^T K_L \int \tanh(s) ds \end{aligned} \quad (16)$$

$$\begin{aligned} \dot{V}(s) &= s^T [\tau - Y_r(q, \dot{q}, \ddot{q}_r, \dot{q}_r) \phi - K_d s] - s^T C(q, \dot{q}) s \\ &+ \frac{1}{2} s^T \dot{D}(q) s + s^T K_L \int \tanh(s) ds \end{aligned} \quad (17)$$

Applying the antisymmetric property in (17), it reduces to the following expression:

$$\dot{V}(s) = s^T \left[\tau - Y_r(q, \dot{q}, \ddot{q}_r, \ddot{q}_r) \phi - K_d s + K_L \int \tanh(s) ds \right] \quad (18)$$

For asymptotic convergence (Lyapunov stability theorem) is required $\dot{V}(s) \leq 0$, in this case a function is designed \dot{V} as follows:

$$\dot{V}(s) = -s^T K_D s \leq 0 \quad (19)$$

$$\dot{V}(s) = -s^T K_D s = s^T \left[\begin{matrix} \tau - Y_r(q, \dot{q}, \ddot{q}_r, \ddot{q}_r) \phi - K_d s \\ + K_L \int \tanh(s) ds \end{matrix} \right] \quad (20)$$

In (20), can be reduced to the following expression:

$$-s^T K_D s = s^T \left[\begin{matrix} \tau - Y_r(q, \dot{q}, \ddot{q}_r, \ddot{q}_r) \phi - K_d s \\ + K_L \int \tanh(s) ds \end{matrix} \right] \quad (21)$$

Solving for τ in (21) is obtained:

$$\tau = Y_r(q, \dot{q}, \ddot{q}_r, \ddot{q}_r) \phi + K_d s - K_D s - K_L \int \tanh(s) ds \quad (22)$$

Where $K_D > k_d$, so that $K_d s - K_D s$ can be written as $-k_{ds} s$, therefore the control law for systems n degrees of freedom can be written as:

$$\tau = Y_r(q, \dot{q}, \ddot{q}_r, \ddot{q}_r) \phi - K_{ds} s - K_L \int \tanh(s) ds \quad (23)$$

Given that $Y_r(q, \dot{q}, \ddot{q}_r, \ddot{q}_r) \phi$ is upper bounded, we propose modified control law as:

$$\tau = -K_{ds} s - K_L \int \tanh(s) ds \quad (24)$$

5 EXPERIMENTAL PLATFORM FOR DIAGNOSIS AND REHABILITATION

The purpose of this paper is to generate the experimental platform, which permit an effective medical rehabilitation in the shortest time possible, also provide information for a medical diagnosis. The experimental platform consists of the patient with physical disability doing recovery exercises, these exercises are based on the solution maze. The desired trajectory is programmed into the device PHANTOM haptic premium 1.0 [11], this device guides the patient on the maze.

The patient is stimulated in a kinesthetic way (related to muscle and tendons) by the haptic device, however the environment with which interacts is real, the labyrinth is made of wood but the desired trajectory is programmed into the device.

In the Fig.1 presents the haptic guidance system, where X_r is the reference position of the desired trajectory, \dot{X}_r determines the reference velocity of the trajectory, previously evaluated under the best conditions for interaction. Because the design control signal was

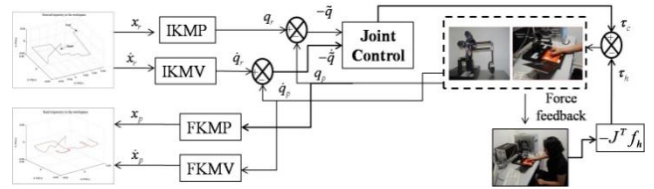


Fig. 1. Haptic guidance system[15]

performed with the joint space required to use models IKMP (Inverse Kinematic Model Position) and IKMV (Inverse Kinematic Model Velocity), so q_r is obtained which represents the vector of desired joint positions and \dot{q}_r is the vector of desired joint velocity, in this way is obtained q which represents the vector of real joints position and \dot{q} represents the vector of real joints velocities. By getting the joint position errors and velocity articulated errors, they are used in the joint control to generate the vector of torques. The patient is manipulated by the haptic device, but the patient disrupts desired tracking trajectory, the perturbation (in position and velocity) is obtained by reading optical encoders mapping is performed after the joint space to the operational space to display the operational positions and velocities.

5.1 Phantom Haptic Device Platform

The experimental platform consists of a PC running at 1.4 GHz with two AMD Athlon processor, 1 Gb RAM, 64 Mb NVIDIA GeForce 4Mx and Visual C++ as a programming language over Windows XP. The haptic device used is PHANTOM premium [11], provide 3 degrees of freedom positional sensing and 3 degrees of freedom force feedback, and connects to the PC via the parallel port (EPP) interface.

5.2 Path Planning For Guidance

The patient is guided in the real world by the end effector of PHANTOM 1.0, in order to solve a maze constructed of wood, the guidance in the maze has a duration is 23.5 seconds. In the Fig. 2 shows the workspace where they are

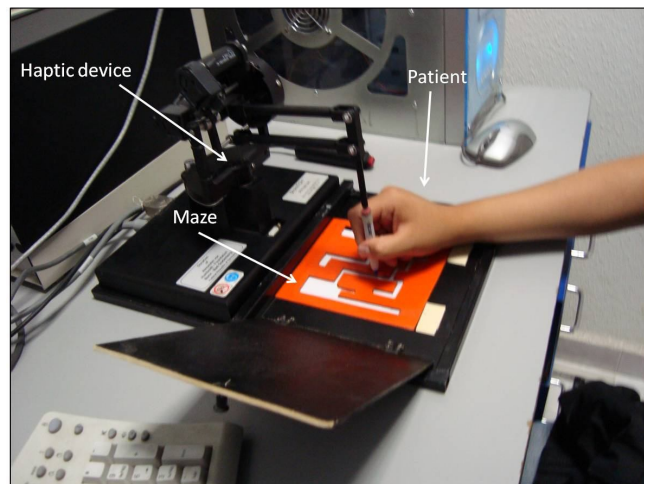


Fig. 2. Experimental platform

held on the rehabilitation of patients.

The polynomials that represent the trajectories of the solution maze (see Fig. 3) are based on equations; they represent the position, velocity and acceleration of the trajectory. The polynomial of the position is as follows,

$$\xi(t) = a_3 \frac{(t-t_0)^3}{(t_b-t_0)^3} - a_4 \frac{(t-t_0)^4}{(t_b-t_0)^4} + a_5 \frac{(t-t_0)^5}{(t_b-t_0)^5} \quad (25)$$

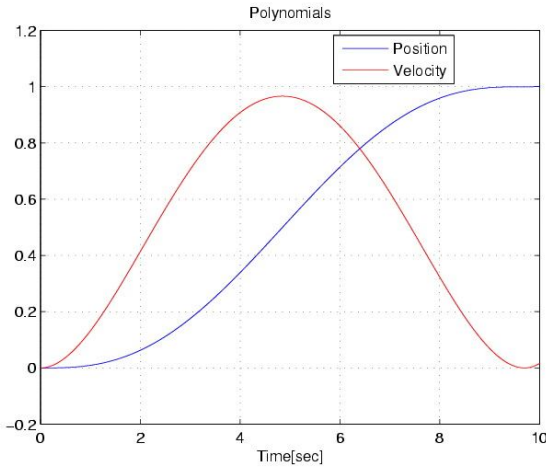


Fig. 3. Performance of polynomials ξ and $\dot{\xi}$.

The polynomial $\xi(t)$ corresponds to a smooth trajectory to the start and end time, which breaks with the inertial effects due to the state of repose and movement (see Fig. 3), t is the time, t_0 represents the initial time and t_b determines at the time of convergence. To obtain the velocity is necessary to derive the polynomial $\dot{\xi}(t)$ with respect to time, and the acceleration has to derive the polynomial twice with respect to time. To determine the values of a_3 ; a_4 and a_5 is done as follows, with the conditions $\xi(t_0) = 0$, $\dot{\xi}(t_0) = 1$, $\xi(t) = 0$, $\dot{\xi}(t_b) = 0$ and $\ddot{\xi}(0.5t_b) = 0$, so solving the equations and is obtained that $a_3 = 10$, $a_4 = 15$ and $a_5 = 6$.

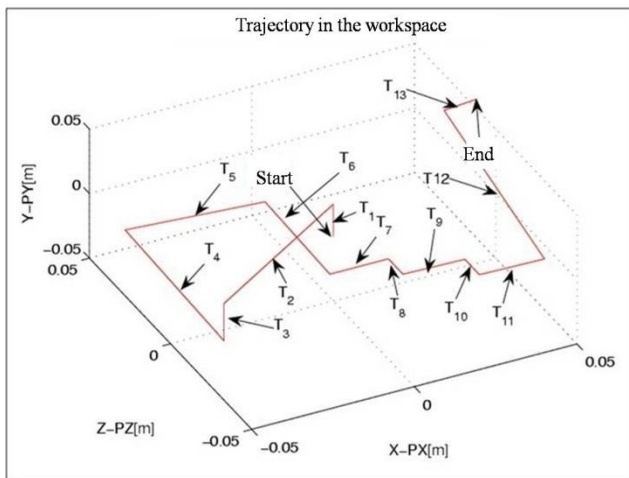


Fig. 4. Path planning

This way we perform the task planning considering convergence in finite time (see Fig. 3). The characteristics that have the polynomials design allows a patient to be guided at the beginning of the exercise: smoothly increasing velocity and decreasing when the trajectory is about to end. In Table I is presented the start and the end time for each trajectory represent in the Fig.4.

TABLE 1
INITIAL AND CONVERGENCE TIME OF THE TASKS PRESENTED IN FIG. 4

Task	Initial time	Finite time convergence
T_1	0 sec.	$t_{b1}=1$ sec.
T_2	1 sec.	$t_{b2}=3$ sec.
T_3	5 sec.	$t_{b3}=1$ sec.
T_4	6 sec.	$t_{b4}=2$ sec.
T_5	8 sec.	$t_{b5}=1.5$ sec.
T_6	9.5 sec.	$t_{b6}=1$ sec.
T_7	10.5 sec.	$t_{b7}=1$ sec.
T_8	11.5 sec.	$t_{b8}=1$ sec.
T_9	12.5 sec.	$t_{b9}=1$ sec.
T_{10}	13.5 sec.	$t_{b10}=1$ sec.
T_{11}	14.5 sec.	$t_{b11}=1$ sec.
T_{12}	15.5 sec.	$t_{b12}=4$ sec.
T_{13}	19.5 sec.	$t_{b13}=3$ sec.

5.3 The Inertial Dynamics Effects

Convergence in finite time by the position and velocity trajectory is reflected in accelerating \ddot{q}_r , this benefits the task planning. Fig.5 shows the inertial torque $\tau_H = D(q)\ddot{q}$ using the patient acceleration generated in the exercise.

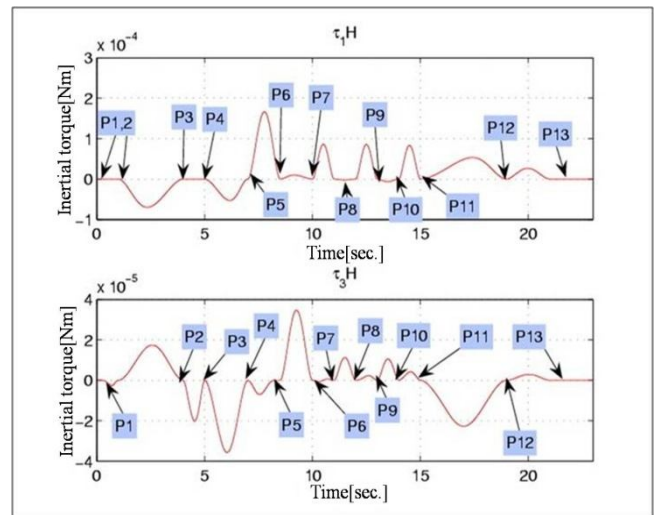


Fig. 5. Inertial torques one and three

However $D(q)$ corresponds to the matrix of inertia of the haptic device. In P0, P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12 and P13 inertia is minimal due to the benefit of the task planning, and the design of the control law that allows stable tracking [16].

6.1 The Experiments

Rehabilitation process consists on patient performance of ten exercises, which have different characteristics under certain parameters. Here are the characteristics of each exercise:

1. Exercise 0.-The patient is guided to follow up the trajectory, in this exercise haptic device actuators are zero, and this way the patient voluntarily moves the PHANToM end effector (see Fig.6).

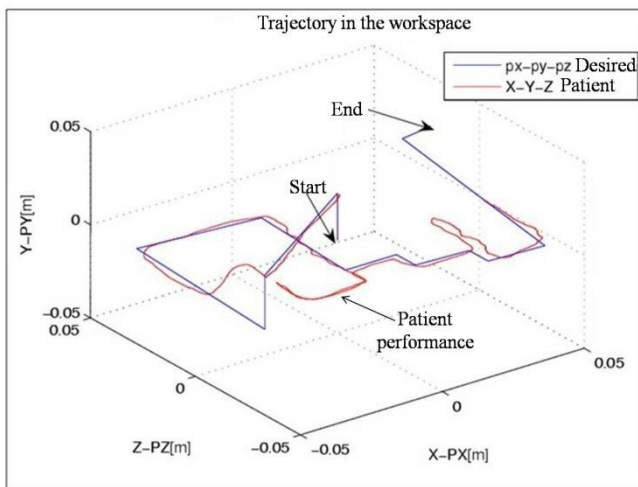


Fig. 6. Exercise 0 (uncontrolled).The patient's performance in the first experiment without control. The patient is not guided by the haptic device PHANToM.

2. Exercise 1-10.-The patient is guided by PHANToM considering the control signal in (24). The constants have the following value $K_{ds} = 0.03$ and $K_L = 0.009$.
3. Exercise 11.-The patient performs the last exercise with the same features as the first, without control in the actuators (see Fig.8). The patient is free movement during the experiment. He controls his movements.

The Fig.7 shows the performance PHANToM 1.0 workspace, the patient has a certain disturbance index during the haptic guidance. Finally, Fig.8 illustrates the performance of the last exercise of patient, this to control the actuators.

6.2 Stability Analysis

The nonlinear control design described in 24 is a robust control that allows a stable man-machine interaction. Control design is based on Lyapunov theory (second method) and passivity. It is well known that one of the

techniques to analyze stability is through the behavior of the Lyapunov function (10) and their respective derivative (20).

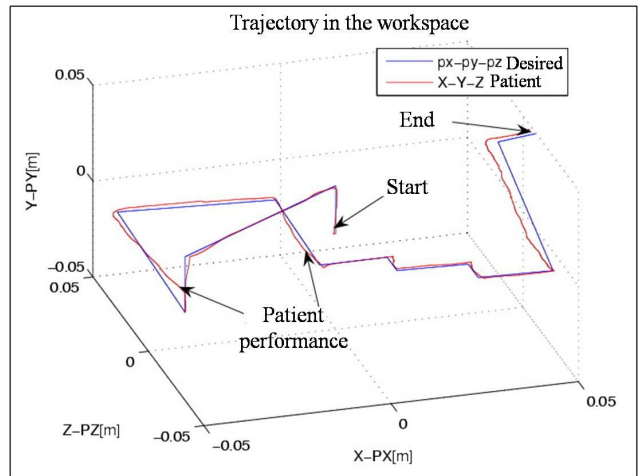


Fig. 7. Exercise 1 (with control) . The patient is guided for 10 times continuously. This allows for success in motor performance

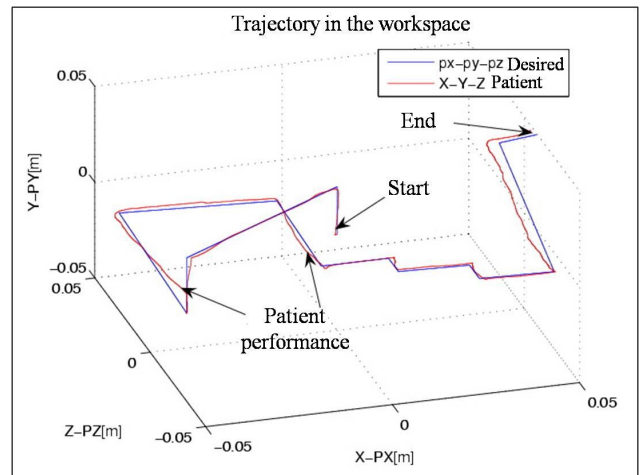


Fig. 8. Exercise 1 (with control) . The patient is guided for 10 times continuously. This allows for success in motor performance

If the candidate Lyapunov function is defined positive in every instant of time and also the derivative with respect to time of the Lyapunov function is negative semi-definite then the system is stable. In order to know if the system (human machine interaction) was stable it is necessary to consider the behavior of the functions described above.

In Fig 9 shows the behavior of the Lyapunov function of Exercise 1. The function is always positive definite, therefore the system is stable with the patient in the control loop.

On the other hand the derivative of the Lyapunov function is shown in Fig 10. The function is negative semidefinite at all times, so that the system is stable at all times. Disturbances of the patient during the interaction are reflected in the impulses that show the functions.

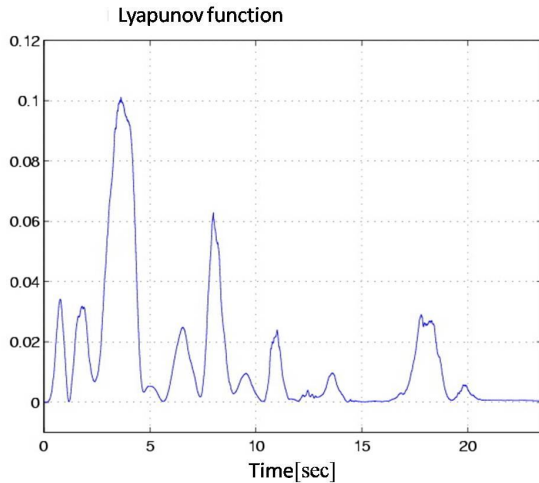


Fig. 9. The graph that describes the behavior of the Lyapunov function permite note that the function was always positive definite, but at certain moments presented impulses, this is due to disturbances of the patient

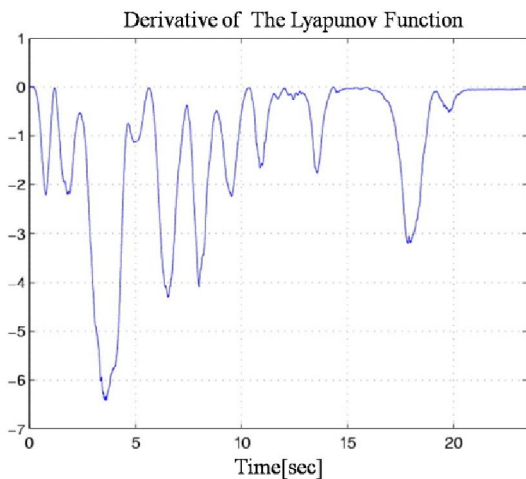


Fig. 10. The high disturbances generated by the patient did not prevent the system was unstable.

Considering the forces on individual fingers, which should be under 30 to 50N in total and the average force that humans can exert on the index finger is 7N, 6N middle finger and ring finger without experiencing 4.5N discomfort or fatigue. It is important that the control signal in the actuators of the haptic device does not exceed the force of 7N. Thus patient safety prevails in closed loop and increasing the useful life of the device. In order to ensure patient safety is necessary to observe the behavior of the nonlinear control law.

In Fig 11 shows the control signal of the three actuators which does not exceed 7N. Because the design of the control law is based on theories that allow for energy analysis and pasivisar the system with the patient in the control loops.

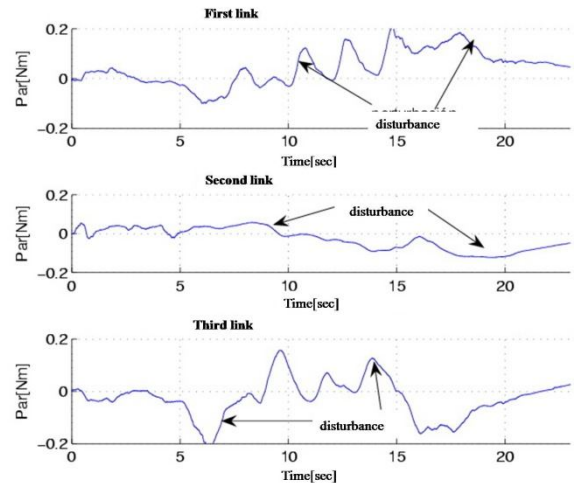


Fig. 11. The interference generated by the patient is reflected in instantaneous changes of the control signal in the three actuators.

6.3 Discussions

In the Fig.12 and Fig.13 present the graphs representing the rehabilitation of the patient during each of the 10 exercises. In each of the experiments $\sum|\tau|$ were performed, measuring the disturbance index that the patient generated during each exercise and then assess the performance and learning via the control signals (see Fig. 9A). The disturbance of the patient influences the dynamic error signals S , in each exercise is calculated $\sum|S|$ and is plotted as shown in Fig. 13.

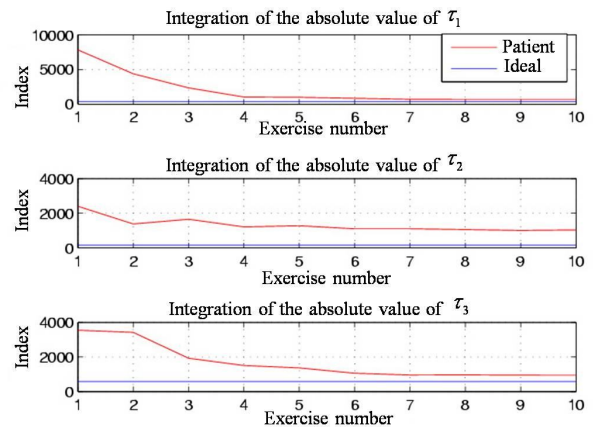


Fig. 12. The performance of the patient in the rehabilitation process is presented in this graphic.

Based on the data obtained, the comparison between the initial results (exercise 1) and the end (exercise 10) results, shows that the patient acquires the ability to solve the maze and also that her motor movements are recovered gradually as the number of exercises is increased.

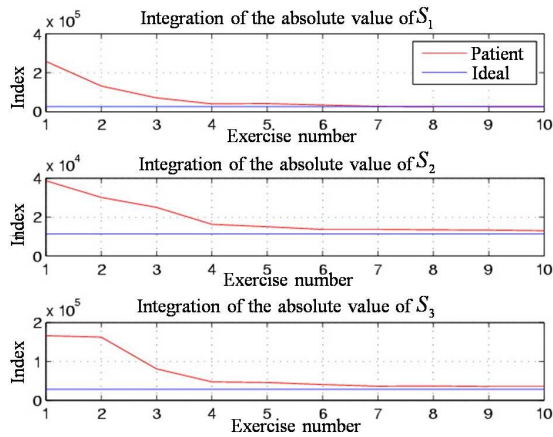


Fig. 13. The absolute value of the error spread in the patient's rehabilitation is important. Reflects the progress of the patient at all

4 CONCLUSIONS

The experimental platform design presented in this work demonstrates the importance of using the technique of haptic guidance for rehabilitation and medical diagnosis. One of the factors that influence the experimental platform for the stability of the system is the design of the control law, which promotes stability and robustness in human-machine interaction, this allows the patient tracking task. This will take care of the patient's physical integration and partial or total damage of the device. Another relevant factor is the high performance that has the haptic device, which in this case is very high due to the high resolution optical encoders as well as the under backlash that presents PHANTOM.

The patient, gradually rehabilitated, using haptic guidance platform leads to reduction in motor rehabilitation time, which reduces the cost that this implies. Data provided by the platform are evaluated by specialists to determine a diagnosis based on the performance of the patient.

REFERENCES

- [1] T. Truelsen, B. Piechowski-Jwiak, R. Bonita, C. Mathers, J. Bogousslavsky, and G. Boysen, "Stroke Incidence and Prevalence in Europe: a Review of Available Data", *European journal of neurology*, vol.13(6), pp. 581-598, June 2006.
- [2] R. Colombo, F. Pisano, S. Micera, A. Mazzone, C. Delconte, M. C. Carrozza, P. Dario, and G. Minuco, "Robotic Techniques for Upper Limb Evaluation and Rehabilitation of Stroke Patients", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol.13, no. 3, pp.313- 324, September 2005.
- [3] P. S. Lum, C. G. Burger, P. C. Shor, M. Majmundar, and M. van der Loos, "Robot-Assisted Movement Training Compared with Conventional Therapy Techniques for the Rehabilitation of Upper Limb Motor Function After Stroke", *Arch. Phys. Meical. Rehabil.* vol. 83, pp. 952-959, 2002.
- [4] H.I. Krebs, J.J. Palazzolo, L. Dipietro, M. Ferraro, J. Krol, K. Ranneklev, B.T. Volpe, and N. Hogan, "Rehabilitation Robotics: Performance-Based Progressive Robot-Assisted Therapy", *Autonomous Robots*, vol. 15, no. 1, pp. 7-20, 2003.
- [5] W.S. Harwin, J. Patton, and V.R. Edgerton, "Challenges and Opportunities for Robot Mediated Neurorehabilitation", *Proceedings of the IEEE Special issue on medical robotics*, vol. 94, no. 9, pp. 1717 1726, September 2006.
- [6] MJ, Eriksson J, Feil-Seifer DJ, Winstein CJ, "Socially assistive robotics for post-stroke rehabilitation", *J Neuroeng Rehabil*, 2007, 4:5.
- [7] P. Metherall, EA. Dymond, N. Gravill, "Posture control using electrical stimulation biofeedback: a pilot study", *J Med Eng Technol* 1996, 20:53-59.
- [8] D. Jack, R. Boian, AS. Merians, M. Tremaine, GC. Burdea, SV. Adamovich, M. Recce, Z. Poizner, "Virtual reality-enhanced stroke rehabilitation", *IEEE Trans Neural Syst Rehabil Eng*, 2001, 9:308-318.
- [9] D. Feygin, M. Keehner and F. Tendick, "Haptic Guidance: Experimental Evaluation of a Haptic Training Method for a Perceptual Motor Skill", *Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2002. HAPTICS 2002, pp 40-47, Orlando, FL, 2002.
- [10] F. Clark and K. Horch, "Kinesthesia: in Handbook of Perception and Human Performance", v.1 *Sensory Processes and Perception*, New York: Wiley, 1986.
- [11] SensAble Technologies, Inc., "PHANTOM PREMIUM 1.0/1.5A, 3D Touch Components": *Hardware Installation and Technical Manual*, Revision 6.5, 18 August,2000.
- [12] V. S. Rafael Kelly. *Control de Movimiento de Robots Manipuladores*. Prentice-Hall, pp 15-37, 2003.
- [13] P. J. N. Romeo Ortega, Antonio Lora and H. Sira-Ramrez. *Passivity-based Control of Euler-Lagrange Systems*. Springer-Verlag, 1998.
- [14] A. Jarillo-Silva and O.A. Domínguez-Ramírez, "Joint Control Strategy for Haptic Guidance", *2010 Electronics, Robotics and Automotive Mechanics Conference*, pp 411-416, 2010.
- [15] A. Jarillo Silva and O.A. Domínguez Ramírez, "Diseño e Implementación de Técnicas de Control No Lineal Basadas en Pasividad Aplicadas a Guiado Háptico", Master, Research Center on Information Technologies and Systems, Autonomous University of Hidalgo State, Hidalgo, México, 2009.
- [16] A. Jarillo-Silva, O.A Domínguez-Ramírez and V. Parra-Vega, "Haptic Training Method for a Therapy on Upper Limb", *3rd International Conference on Biomedical Engineering and Informatics (BMEI 2010)*, pp. 1750-1754, 2010.

Prevalence of Eating Disorders among Female Students of Tonekabon University (Tonekabon-Iran)

Roxanna Ast , Azade Fadavi Roodsari

Abstract—aim of this research is to study the prevalence of eating disorder amongst female students of Tonekabon University. The community being studied is the female students of Tonekabon University. 300 students were randomly selected and requested to complete the 'eating attitude test-26'.

Index Terms— Eating disorders, Female students, Tonekabon University

1 INTRODUCTION

Eating disorder presents a serious problem these days. Every year the number of people suffering from anorexia and bulimia increase and consequents of the disease may be health or even life threatening.[1]

This disorder presents a significant problem among adolescent and young women in many westernized countries and is associated with nervous, physical and psychiatric problems.[2]

Anorexia nervosa is a psychological and physical condition of semi starvation in which individuals weigh 85% or less of what would ordinary be there healthy body weight, resulting in physical impairments and in the 90% of patients who are females Cessation of menses.

This condition is to due to highly restricted food intake, often accompanied by excessive exercise and sometimes purging by self-induced vomiting, laxative use or other means.

These behaviors are usually related to obsessional and a perfectionist thinking that focuses on a distorted body image and fear of becoming fat.[3]

Feeling normal fullness after eating is felt as discomfort and experienced as a failure of control, moral weakness and a source of great guilt. These perfectionists are failing in their major (anorexia) project and will redouble their effort by eating nothing for a day or even restricting their water intake. This can precipitate death through cardiac arrest, particularly if they are also vomiting the little intake they do allow.[4]

The association between anorexia nervosa and depression has long been recognized by clinicians.[5]

A study in Sweden using structured interview for DSM-III-R criteria found that 85% of patients with anorexia nervosa (AN) had a depressive disorder.[6]

About half of sufferers eventually develop binge-eating episodes – that is, periodic decontrol over eating on incapacity to satiate. [7]

Bulimia nervosa (BN) is a condition in which individuals binge, eat large calories, up to 2000 at a time or more and

then purge themselves of what they have eaten usually by forcing themselves to vomit and sometimes by means of laxatives, diet pills, diuretic pills or excessive exercising.

These behaviors occur at least several times per week for months on end, the condition is usually related to over concern with ones weight and shape, and is accompanied by feeling of shame, disgust and being out of control.[8]

Like anorexia nervosa, Bulimia was recognized to occur as early as the 17th century. [9]

According to Wilson [10] to be diagnosed with bulimia nervosa individuals must experience episodes of binge-eating 'at least twice a week' on average, for three to six months.

In addition to the primary eating disorders, several other conditions occur among individuals with psychiatric disorders that may markedly affect eating behavior and weight. For instance, individuals with severe depression experience an increase in appetite and food cravings. Patients with psychic delusions due to Schizophrenia or other conditions may think food is poisoned and refuse to eat.[11]

Recent studies have found that EDNOS (Eating Disorder Not Otherwise Specified) is the most common eating disorder, diagnosis both in outpatients and inpatients settings. Underweight patients that do not report over-evaluation of shape and weight are a distinctive and scarcely studied subgroup of EDNOS.

Their self-evaluation is largely or exclusively based on their abilities to control their eating purge.[12]

The incidence of eating disorders in females has been extensively studied in both anorexia nervosa and bulimia nervosa.[13]

In a study of 105 patients with eating disorders, Braun [14] found that the life time prevalence of any offensive disorder was 41.2 % in anorectic restrictors, 82% in anorectic bulimics, 64.5% in patients with bulimia nervosa and 78% in patients with bulimia nervosa with a past history of anorexia nervosa.

According to Treasure [15] the present time prevalence of

all eating disorders is about 5%.cultural, social and inter-personal elements can trigger onset and change in networks can sustain the illness.

Although it is clear that anorexia nervosa occurs in men as well as women, and in younger as well as in older people, few studies report incidence rate for males or for people beyond the age of 35.

The majority of male incidence rate reported was below 0.5 per 100,000 populations per year. [16]

Studies have reported the female to male ratio to be around 11 to 1.[17]

On an overall female rate of 15.0 per 100,000 population per year, Lucas [18] reported a rate of 9.5 for 30-39-year-old women,5.9 for 40-49-year-old women,1.8 for 50-59-year-old and 0.0 for women aged 60 and over .

According to a research by Casper [19], women who had recovered from anorexia nervosa rated higher on risk avoidance, displayed greater restraint in emotional expression and initiative, and showed greater conformance to authority than age-matched normal women.

Lucas [20] found that the age-adjusted individuals rates of AN in females 15-24-year-old showed a highly significant linear increasing trend from 1935-1989,with an estimated rate of increase of 1.03 per 100.000 person per calendar year.

2 METHOD

This is a descriptive study and aims to objectively describe the features of the desired community.

The research took place over a period of 4 months. (March 2010 - June 2010).

In this research, based on the total population -6000 female students-a sample of 300 were randomly selected. Subjects were students who attended classes at university throughout the week. The team, attended classes on different hours of different days and selected few names from the participants of the class-regardless of their BMI, body shape and weight-afterwards, subjects completed the questionnaire. EAT-26 was used as the main method of data collection.

The eating attitude test (EAT-26) has been proposed as an objective, self report, measuring the symptoms of anorexia nervosa. It has been used as a screening instrument for detecting previously undiagnosed cases of anorexia nervosa in populations at high risk of the disorder. [21]

Subjects can score between 0 – 78.any score above 20 is diagnosed with eating disorder.

The EAT-26 can only show some symptoms of eating disorder. To classify the type of disorder (AN or BN) a diagnostic interview will need to take place.

3 RESULTS

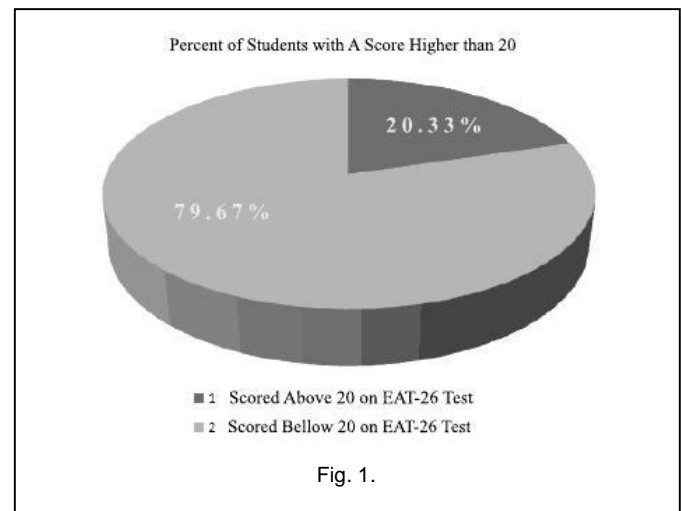
The results of this study showed that among 300 cases under review, 61 of them received a score higher than 20 on the EAT-26 test and were diagnosed with eating dis-

orders.

TABLE 1. EAT-26 TEST RESULTS

SAMPLE POP-ULATION	STUDENTS SCORING ABOVE 20	PERCENTAGE OF STUDENTS WITH A SCORE HIGHER THAN 20
300	61 STUDENTS	20.33 %

This amount forms 20.33 % of the total population of female students at Azad university of Tonekabon.



Among the three sub-scales namely: Oral Control – food habits and Esurience (desire to eat), the most points belonged to food habit sub-scale (86.88%) second is oral control and the rest were high in esurience sub-scale.

The main problem and barrier throughout the research was the subject's orientation towards the questions on the test which was resolved by the researcher's explanation in person.

4 DISCUSSION

The purpose of this study is to measure the prevalence of eating disorder in female students of Tonekabon University. According to the results 20.33% of subjects were identified with eating disorder.

In our country, in the field of eating disorder, few studies have been carried out.[22]

In measuring the prevalence of eating disorders amongst high school students (male and female) in Sari- North of Iran-In the academic year of 2002-2003 in 10.5% of the students, abnormal attitude towards eating was observed. This conclusion suggests the frequency of abnormal attitude toward eating disorder is more or less similar to that

of other communities.

Patients with this attitude most likely have some psychological disorder such as, depression. If this issue is not managed in the appropriate manner, many side effects will follow. [23]

AN and BN occur disproportionately often in industrialized (vs. economically less developed countries) nations, but are otherwise reported to occur with surprisingly uniform prevalence in developed parts of Europe, Asia and the Americas. [24]

Young girls and women have been influenced by western culture and have adapted the same behaviour towards issues like appearance and body shape. [25]

The extent of the impact has not, yet, been provided in statistics. [26]

In a study to determine the prevalence of eating disorder among female high school students in Tehran-Capital of Iran-results show that 9.0 % were diagnosed with anorexia, 23.3% with Bulimia and 63.7% with a mild eating disorder. [27]

In another study by Seyedi [28] the prevalence of eating disorder among 600 female high school students in Kerman was found to be 7.6%.

Some studies suggest an increase in incidence of AN over various time period, whereas others report an increase in incidence followed by stabilization. [29]

Epidemiological studies show that eating disorder is mostly a white female disease in industrialized countries. Societies, in which, slimness is a main criteria of attraction. [30]

Unfortunately, in the past few years eating disorders has increased in Iran. As the most people diagnosed with this disorder are the young generation, therefore, education on this matter is one of the most important means of handling this issue.

Generally, the level of research conducted on aetiology and treatment of eating disorder In Iran is very low. Young girls of today will be the mother of our future generation and their mental and physical health plays an important role in the future of our children and country. Therefore, this field need far more extensive research and exact aetiology and strategies for prevention and therapy.

5 CONCLUSION

The results show that, based on the eat-26 test criteria, from the total population under study, 61 students, and equal to 20.33% of the total number of students- were diagnosed with eating disorders. Eating disorders are breaking out rapidly in our society and among the young generation and if not dealt with properly serious consequences will be expected.

REFERENCES:

- [1] Pasternak B., Czekaj J., Sabina level of anxiety connected with putting on weight as a predictor of bulimia and anorexia.,2009 , physiothera vol.17,issue.3, p36 5p.
[2] Brown M., Fary Cachelin., Faith A., eating disorder in

ethnic minority women: a review of the emerging literature, Current psychiatry review,2009, vol.5,issue.3, p 182-193

[3],[8],[11] Yager J., Encyclopaedia of psychotherapy, eating disorder, 2002, p671, 672

[4] Churven P., Hunger strike or medical disorder? Is anorexia caught in our flawed dichotomy between body and mind?, ANZJFT , 2008 , Vol.29, Num.4 , pp.184-190

[5],[7],[9],[24] Maj M., Halmi K., Lopes J.J, Sartorius N., Eating disorder, 2003, vol.6 , ISBN :0-470-84865-0

[6] Ivarsson T., Rastam M., Wentz E., Gillberg I., Gillberg.C ., depressive disorders in teenagers onset anorexia nervosa : a controlled longitudinal , party community-based study .(2000),compr, Psychiatry , 41 :398-403

[10] Wilson G.T., Sysko R., Frequency of binge eating episodes in bulimia nervosa and binge-eating disorder: diagnostic consideration, International journal of eating Disorder, Nov 2009, 42(7): 603-10 ISSN: 0276-3478

[12] Grave R., Calugi S., Marchesini., Underweight eating disorder without over-evaluation of shape and weight: atypical anorexia nervosa?, International journal of eating disorders, 2008, 41: 8705-712

[13],[29] Button E., Aldridge S., Palmer R., International Journal of eating disorders, males assessed by a specialized adult eating disorders service: patterns over time and comparisons with females ,41:8 ,758-761, 2008

[14] Braun D.L ., Sunday S.R., Halmi K.A., psychiatric comorbidity in patients with eating disorders, 1994 , psycho , med ,24 : 859-867

[15] Treasure J.,Claudino AM., Zucker N Lancet, Eating disorder,13 Feb. 2010, vol. 375 pp 583-593,England

[16] Turnbull L.S., Ward A., Treasure J., Jick H., Derby L., The demand for eating disorder care, an epidemiological study using the general practice research data base, 1996 , British journal of psychiatry ,169 , pp 705-712

[17] Hoek H.W., Bartelds , A.I.M ., Boosveld J.J.F, Vander Graaf, Limpense Y., V.E.L ., Maidwald M & Spaaij C.J.K , Impact of urbanization on detection rates of eating disorders, 1995, American journal of psychiatry ,152, pp 1272-1278

[18],[20] Lucas A.R, Crowson C.S, O'Fallen , W.M & Meltroy L.J, The ups and downs of anorexia nervosa , 1999 , international journal of eating disorders , 26 , 397-405

[19] CASPER RC., personality features of women with good outcome from restricting anorexia nervosa, psychosomatic medicine, 1990, Vol.52, Issue.2, p156-170,

[21] The eating attitude test, Psychometric features and clinical correlates, psychological medicine, 1982, 12, p 871-878

[22] Babae S., Relationship between core beliefs and ability to eat with symptoms of eating disorder in students of shahid beheshti University, General psychology Senior thesis , 2006.

[23] Zarghami M.,Chyme N., Prevalence of non-natural approach to eating and the factors correlated with the high school students,2002, Journal of Mazandaran Uni-

versity of medical sciences, Autumn 2003 ; 13 (40) :70-78

[25] Gaffari Nejd A., eating disorders in female students of kerman university of medical sciences, ministry of health and medical education, Kerman University of medical sciences, 1998

[26],[28] Seyedi F., Sanayi F., Prevalence of eating disorders in female students in high schools of Kerman, 2004, Kerman university school of medical sciences.

[27] Nobakht M., epidemiology study of eating disorders in second year high school female students in Tehran, Thesis (MA) 1998, Ferdowsi University of mashhad, Faculty of educational sciences and psychology

[30] Szabo C.P. , Hollands. C ; Abnormal eating attitudes in secondary-school girls in south Africa; South Africa medical journal , 1997 , 87(4) , 524- 6: 528-30

Impact of Leverage on Firms Investment Decision

Franklin John. S, Muthusamy. K

Abstract - The present paper is aimed at analyzing the impact of leverage on firm's investment decision of Indian pharmaceutical companies during the period from 1998 to 2009. To measure the impact of leverage on firm's investment decision, pooling regression, random and fixed effect models are used by taking, leverage, sales, cash flow, Return on Asset, Tobin's Q, liquidity and retained earnings as independent variable and investment as dependent variable. In addition, we demarcate between three types of firms (i) Small firms, (ii) Medium firms and (iii) Large firms. The results reveal that a significant positive relationship between leverage and investment, while we found a negative relationship between leverage investment for medium firms and positive relationship between leverage and investment in large firms. Our econometric results reveal an insignificant relationship between the two variables for medium and large firms.

Index Terms-- Investment, Tobin's Q, Cash flow, Liquidity, ROA, Size and Retained Earnings.

INTRODUCTION

Investment is a crucial economic activity in the corporate financial management. Such an activity leads to the country's economic development provide employment to the people and to eliminate poverty. This paper investigates the effort of debt financing on the firms investment decision on pharmaceutical industry in India. It plays a significant role in the country's economic and industrial development and trade and to prevent diseases' for increasing the life of people. This industry is providing a basic material to other industrial sectors. It requires capital for financing firm's assets. Among the different sources of fund, debt is a cheaper source because of its lowest cost of capital. The investment decision of the firm is of three categories that can be adopted by firm's management besides the financing decision and the net profit allocation decision. The investment

decision has a direct influence on the firms asset structure, more over in their degree of liquidity and consists of spending the financial funds for the purchase of real and financial assets for the firm. In order to gain cash and the growth of the wealth of firms owner. The investment decision and the financing decision are interdependent that is the investment decision is adopted in relation to the level of financing source but the option to invest is also crucial in order to calculate the level of financing capitals and the need for finding their sources.

As far as the hierarchy of financing sources as it exists in the economic literature, is concerned, cash flow is the cheapest financing sources followed by debts and in the end, by its issuing of new shares. Debts can be cheaper than the issue of new shares because the loan contract can be created as to minimize the consequences of

information problem. Giving the fact the degree of information asymmetry and the agent costs depend on the peculiarities of every firm, such firms are more sensitive to financial factors than other. The debt limit of the firms is determined in the view, since interest payment is tax deductible, the firm prefers debt financing to equity and it would rather have an infinite amount of debt, However, this leads to negative equity value in some status so that the firm would rather go bankrupt instead of paying its debt. Therefore debt to remain risk-free, lenders will limit the amount of debt. They can limit the debt by accepting the resale value of capital as collateral and ensuring that this value is not lower than the amount of debt, so that they can recover their money in case of bankruptcy. Alternatively, lenders may limit the amount of debt in order to ensure that the market value of equity is always non-negative and bankruptcy is sub-optimal for the firm.

While there is by now a rapidly expanding literature on the presence of finance constraints on investment decisions of firms for developed countries, a limited empirical research has been forthcoming in the context of developing countries for two main reasons. First until recently, the corporate sector in emerging markets encountered several constraints in accessing equity and debt markets. As a consequence, any research on the interface between capital structure of firms and finance constraints could have been largely constraint-driven and have less illuminating.

Second, several emerging economies, even until the late 1980s, suffered from financial depression, with negative real rates of interest as well as high levels of statutory pre-emption. This could have meant a restricted play of market force for resource allocating.

Issues regarding the interaction between financing constraint and corporate finance have, however, gained prominence in recent years, especially in the context of the fast changing institutional framework in these countries. Several emerging economies have introduced market-oriented reforms in the financial sector. More importantly the institutional set-up within which corporate houses operated in the regulated era has undergone substantial transformation since the 1990s. The moves towards market-driven allocation of resources, coupled with the widening and deepening of financial market, have provided greater scope for corporate house to determine their capital structure.

The rest of the paper unfolds as follows. Section II discusses the historical background of the study. Section III explains methodology, data, variable description and the data employed in the paper. Section IV presents the results and discusses robustness check followed by the concluding remarks in the final section.

THE BACKGROUND OF THE STUDY

Several authors have studied the impact of financial leverage on investment. They reached

conflicting conclusions using various approaches. When we talk about investment, it is important to differentiate between overinvestment and underinvestment. Modigliani and Miller (1958) argued that the investment policy of a firm should be based only on those factors that would increase the profitability, cash flow or net worth of a firm. Many empirical literatures have challenged the leverage irrelevance theorem of Modigliani and Miller. The irrelevance proposition of Modigliani and Miller will be valid only if the perfect market assumptions underlying their analysis are satisfied. However the corporate world is characterized by various market imperfections costs, institution restrictions and asymmetric information. The interaction between management, shareholders and debt holders will generate frictions due to agency problems and that may result to underinvestment or over-investment incentives. As stated earlier one of the main issues in corporate finance is whether financial leverage has any effects on investments policies.

Myers (1977), high leverage overhang reduces the incentives of the shareholder-management coalition in control of the firm to invest in positive net present value of investment opportunities, since the benefits accrue to the bondholders rather than the shareholders thus, highly levered firm are less likely to exploit valuable growth opportunities as compared to firm with low levels of leverage a related under investment theory centers on a liquidity affect in that firm with large debt

commitments invest less no matter what their growth opportunities. Theoretically, even if leverage creates potential underinvestment incentives, the effect could be reduced by the firm corrective measures. Ultimately, leverage is lowered if future growth opportunities are recognized sufficiently early.

Another problem which has received much attention is the overinvestment theory. It can be explained as investment expenditure beyond that requires to maintain assets in place and to finance expected new investment in positive NPV projects where there is a conflict between manager and share holder. managers perceive an opportunities to expand the business even if the management under taking poor projects and reducing shareholders welfare. The managers' abilities' to carry such a policy are restrained by the availability of cash flow and further tightened by the financing of debt. Hence, leverage is one mechanism for overcoming the overinvestment problem suggesting a negative relationship between debt and investment for firm with low growth opportunities. Does debt financing induce firms to make over-investment or under-investment? The issuance of debt commits a firm to pay cash as interest and principal. Managers are forced to service such commitments. too much debt also is not considered to be good as it may lead to financial distress and agency problems.

Hite (1977) demonstrates a positive relationship because given the level of financial

leverage an investment increase would lower financial risk and hence the cost of bond financing. In contrast Deangels and Masulis (1980) claim a negative relationship since the tax benefit of debt would compete with the tax benefit of capital investment. Dotan and Ravid (1988) also show a negative relationship because investment increase would raise financial risk and hence the cost of bond financing how the investment increase affects financial risk and the substitutability between tax shields and hence; financial leverage may depend on firm-specific factors.

Jensen (1986) points out that liabilities can help avoid overinvestment by reducing the cash flow left up to corporate manager's own discretion and constraining investment in investment projects that might be desirable for corporate managers but not desirable for companies' future profitability. Jensen argues that whether liabilities restrain overinvestment depends largely on whether companies have growth opportunities. In short, Jensen points out those liabilities have not only the negative effects of restraining overinvestment by low-growth companies. Like Jensen(1986),Stulz(1990) and Hart and Moore(1995) argue that liabilities effectively restrain overinvestment. They reason that increased liabilities, by enlarging repayment obligations, not only curtail free cash flow but also raise the possibility of corporate bankruptcies, thus prompting corporate managers to reduce

investment and sell off unprofitable business divisions.

Daddon and Senbets (1988) hypothesis on the relationship between bond financing and capital investment which is conditional on from specific variables such as tax shield, retention ability, capital intensity and insider equity ownership. Josephic Kang(1995)who found that the level of bond financing has negative relationship with level of investment.

Whited (1992) has shown how investment is more sensitive to cash flow in firms with high leverages as compared to firms with low leverage. Cantor (1990) showed that investment is more sensitive to earnings for highly levered firms

McConnell and Servaes (1995) have examined a large sample of non financial United State firms for the years 1976, 1986 and 1988. They showed that for high growth firms the relationship between corporate value and leverage is negatively correlated. Also the allocation of equity ownership between corporate insiders and other types of investors is more important in low growth than in high growth firms.

McConnell and Servaes (1995) use cross-sectional data to analyze U.S listed companies in 1976, 1986 and 1988, and find "two faces of debt," meaning that enterprise value was negatively correlated with the debt ratio of companies with high growth opportunities. Lang et al. (1996), based on an analysis of the relationship between the debt ratio

and the rate of growth of companies, point out that for companies with fewer investment opportunities (i.e. companies with a low Tobin's Q), there is a negative correlation between the debt ratio and the investment. The estimation results from their studies do not find a negative correlation between the debt ratio and the growth rate for companies with abundant growth opportunities. In other words, for companies with investment opportunities, increased liabilities do not necessarily hamper growth.

Lang et al (1996) found that there is negative relation between leverage and future growth at the firm level and for diversified firms, at the business segment level. Also debt financing does not reduce growth for firms' known to have good investment opportunities, but it is negatively related to the growth for firms whose growth is not recognized by the capital market.

Myers (1997) has examined possible difficulties that firms may face in raising finance to materializing positive net present value (NPV) projects, if they are highly geared. Therefore, high leverage may result is liquidity problem and can affect a firm's ability to finance growth. Under this situation, debt overhang can contribute to the under-investment problem of debt financing. That is for firms with growth opportunities debt have a negative impact on the value of the firm.

Ahn et al.(2000), found that diversified companies tend to have higher debt ratio than

focused counterparts and diversified companies make larger investments (net cost of capital/sales) than focused counterparts. They also point out that debt ratio influence management decisions on investment and that diversified companies can overcome debt ratios through the distribution of liabilities by corporate managers.

Arikawa et al,(2003) adopt the method of estimation used by Lang et al.(1996) and point out that the main bank system in Japan helped amplify the disciplinary function of liabilities, particularly for low-growth companies.

Aivazian et al (2005) analyses the impact of leverage on investment on Canadian industrial companies cover the period from 1982 to 1999. They found a negative relationship between investment and leverage and that the relationship is higher for low growth firms rather than high growth firms.

Ahn et al (2006) found that diversified companies tend to have higher debt ratios then focused counter parts and diversified companies make larger investments than focused counter parts. They also point out that debt ratio influence management decisions on investments and that diversified companies can overcome the constraints of debt ratio through the distribution of liabilities by corporate managers.

Mohanprasadsing Odit and Chitto (2008) analyze the impact of leverage on firms' investment on 27 maturation firms that are quoted on the stock

exchange Mauritian for the year 1990 – 04. They found that leverage has a significant negative effect on investments, Suggesting that capital structure plays an important role in the firms investment policies while the negative relationship persist for low growth firm, this is not the case for high growth firm.

Thus the previous studies have verified the impact of leverage on firm's investment decision as well as the effect of leverage in restraining over investment and facilitating under investment. These studies suggest that leverage restraining over investment but likely cost under investment. Thus in this paper an attempt is made to more clearly the leverage impact on firms investment decision on pharmaceutical companies in India.

METHODOLOGY DATA AND VARIABLES DESCRIPTION

We estimate a reduced form of investment equation to examine the effect of leverage on investment the specification is similar to Aivazian Ge and Qiu (2005). This is as follows:

$$I_{i,t}/K_{i,t-1} = \alpha + \beta [CF_{i,t}/K_{i,t-1}] + \beta_1 Q_{i,t-1} + \beta_2 LEV_{i,t-1} + \beta_3 SALE_{i,t-1} + \beta_4 ROA_{i,t-1} + \beta_5 LIQ_{i,t-1} + \beta_6 RETES_{i,t-1} + \mu_{i,t}$$

Where $I_{i,t}$ represents the net investment of firm i during the period t ; $K_{i,t-1}$ is the net fixed asset; $CF_{i,t}$ is the cash flow of firm i time t ; $Q_{i,t-1}$ is the Tobin's Q ; $LEV_{i,t-1}$ represents the leverage; $SALE_{i,t-1}$ stands for net sales of firm i ; $ROA_{i,t-1}$ is

the profitability of the firm i ; $LIQ_{i,t-1}$ represents liquidity of firm i ; $RETES_{i,t-1}$ is the retained earnings of the firm i .

The data used in this paper are from the annual report of Indian pharmaceutical companies which are listed in Bombay stock exchange and this data have been collected from CIME browse data base of top 25 companies based on sales from the period 1998 – 2009

LEV: Lev denotes leverage. We have used the same definition of leverage as lang. et al (1996), namely the ratio of total liabilities to the book value of total assets. He pointed that the market value leverage gives too much weight to the deviations in equity value. The book value of leverage does not reflect recent deviation in the market valuation of the firm. If leverage has a significant negative effect on investment, two interpretations can be adopted. First, it would mean that capital structure plays an important role in the firm's investment policies; second, it can also be explained by an agency problem between the agents and the shareholders. If managers are overburdened by debt they may give up projects which may yield positive net present values. Also there will be support for both the underinvestment and overinvestment theory.

TOBIN'S Q : we use prefect and wiles (1994) simple Q (market value + liabilities / book value of assets) as a proxy for growth opportunities defined as the market value of total assets of the firm divided by the book value of assets. Market value of the firm

is the sum of total liabilities, the value of equity shares and the estimated value of preference shares. The market value of preference share is calculated as preference divided multiply by ten which measures growth opportunities and it, compare the value of a company given by financial market with the value of a company, Tobin's q would be 1.0 if Tobin q is greater than 1.0 then the market value is greater than the value of the companies record assets. This suggests that the market value reflects some unmeasured or unrecorded assets of the company. High Tobin's q values encourage companies to invest more in capital because they are "worth" more than the price they paid for them. On the other hand, if Tobin's q is less than 1, the market value is less than the recorded value of the assets of the company.

SALE: sale is measured as net sales deflated by net fixed assets. Which measures the efficiency with net fixed assets is measured. A high ratio indicates a high degree of efficiency in asset utilization and a low ratio reflects inefficient use of assets.

CASH FLOW: Cash flow is measured as the total of earning before extraordinary items and depreciation and is an important determinant for growth opportunities. If firms have enough cash inflows it can be utilized in investing activities. It also provides evidence that investment is related to the availability of internal funds. Cash flow may be termed as the amount of money in excess of that needed to finance all positive net present value of

projects. The purpose of allocating money to project is to generate a cash inflow in the future, significantly greater than the amount invested. That is the objective of investment is to create shareholders wealth. In order to eliminate any size effect. We normalize this measure by taking the book value of assets; this method was utilized by Lehn and Poulson (1989) and Lan et al (1991).

PROFITABILITY (ROA): Profitability is measured in terms of the relationship between net profits and assets. It is calculated as earning after tax adds interest minus tax advantage on interest divided total fixed assets. It shows the operating efficiency of the total funds over investment of a firm and is another important variable that are utilized to growth opportunities as it tries to explain how much the assets that the firm is employing in contributing to the total profitability. Therefore investment in assets contributes to the profitability and we can proxy high profitability with high growth firms.

LIQUIDITY (LIQ): the liquidity ratio is measured by the current assets divided by the current liability and is the ability of firms to meet its current obligations. Firms should ensure that they do not suffer from lack of liquidity as this may result in to a sate of financial distress ultimately leading to bankruptcy. Lack of liquidity can lead to a struggle in term of current obligations, which can affect firm's credit worthiness, Bernake and Gerler (1990) argued that "both the quantity of investment spending and its expected return will

be sensitive to the credit worthiness of borrowers". That leads us to say that investment decisions of firms are sensitive to current liquidity. However, Firms with high liquidity give the signal that funds are tied up in the current assets.

RETAINED EARNINGS (RETES): represents the amount of business savings meant for bloughing back. These are the most favored sources of finance for corporate firms. There is a significant difference in the use of internally generated funds by the highly profitable corporate relative to the low profitable firms

RESULT AND DISCUSSION:

This section portrays the result from the regression estimation, we present result for the small size, medium size and larger sized firm is classified based on the size. The smaller size is obtained by subtracting mean from standard deviation of total asset and larger size is obtained by adding mean value of asset to standard deviation. The median sized firms are those firms which are not belong to both categories of the firm. The econometric result for the sample firms is showed the pooled estimates; random effect estimates and fixed effect estimates on the T values are shown in the parenthesis. Two statistics are used in order to identify, which methodology is appropriate to establish the relationship between leverage and investment. First we compare the pooled estimates and random effect estimates. The second Lagrangian Multiplier (LM) test is

performed with a large chi-square values indicate of low P-value. We reject that the pooled estimate is appropriate. The second to compare random effect estimate with fixed effect estimate, the Hausman test is performed. If the model is correctly specified and it the effect are uncorrelated with independent variables the fixed effect and random effect should not be different a high chi-square value is indicate of appropriateness of the fixed effect.

RESULT OF SMALL FIRMS'

Table 1 brings out the regression result of small firms. It shows that the leverage has a positive impact on investment at the 5% significant level. The impacts of other variables on have the expected signs. The retained earnings have a significant positive impact on investment. To identify which empirical methodology pooling random effect or fixed effect regression is most suitable, we perform two statistical test the first the Lagrangian Multiplier (LM) test of the random effect model. The null hypothesis is that individual effect u_i is 0. The chi-square value is 25.74 thus the null hypothesis is rejected at 1% level of significance. The results suggest that the rho effect is not zero and the pooling regression is not suitable in this case the regression co-efficient leverage on small firms from the pooling regression is equal to 1.3451 and is not significant. The regression co-efficient of leverage of firms from

Random effect and fixed effect model are 3.4868 and 1.8200 respectively. The regression co-efficient from the pooling regression are much smaller than those estimated from the random

effect and fixed effect models suggesting that ignoring individual firm effects leads to an underestimation of the impact of leverage on investment.

Table 1:-Regression result of small firms

Variables	Pooling	Fixed effect	Random effect	Fe with AR
Constant	1.3085 (3.40)	2.5918 (6.75)	1.6940 (4.63)	1.2473 (3.14)
leverage	1.3451 (-0.84)	3.4868 (2.65)	1.8200 (1.28)	0.5834 (0.38)
Sales	-0.0002 (-0.42)	-0.0001 (-0.18)	-0.0001 (-0.18)	-0.0003 (-0.51)
ROA	0.0003 (0.18)	0.0002 (0.15)	0.0003 (0.21)	0.0007 (0.32)
Cash Flow	0.2081 (0.46)	0.2264 (0.69)	0.0727 (0.20)	0.2647 (0.52)
Liquidity	-0.1025 (-0.74)	-0.01667 (-0.16)	-0.0724 (-0.60)	-0.0338 (-0.25)
Retained earnings	0.0027 (2.26)	0.0020 (-1.55)	0.0010 (0.84)	0.0027 (2.06)
Tobin's Q	1.1484 (1.52)	-1.0169 (-1.25)	0.6310 (0.85)	1.3154 (1.72)
LM Test	Chi ² (1)=25.74			
Hasuman Test	Chi ² (7)=29.85			
Observations	60	60	60	60
Groups	-	10	10	6
Adjusted R ²	0.2343			

We conduct the Hausman specification test to compare the fixed effect and the random effect models. If the model is correctly specified and of individual effects are uncorrelated with independent variables the statistics are showed that the null

hypothesis is rejected at the 1% significance level. The results suggest that the fixed effect model is most appropriate in estimating the investment equation.

Leverage is statistically significant at 1% and 5% level of significant and is positively related to

investment. A 1 unit increase in the leverage leads to an increase by 3.4568 units in investments this implies that a leverage increases in small firms is also increase a investment of firms because firms do not have a adequate asset cushion for financing the projects. Thus, in a small sized firm tend to be more dependent on debt as a source of finance to finance the projects.

The table also reveals that small firms are under utilizing their fixed assets and it would affect the ability in generating the volume of sales and the co-efficient value is -0.001 and it is not statistically significant.

The co-efficient value of ROA is 0.0003 and is not statistically significant but positively related with investment. It indicates the operating efficiency of the employed funds over investment is positive. Higher the ROA is also attracting funds from investors for expansion and growth.

Cash flow and retained earnings are positively related with investments not statistically significant and coefficient value is 0.2264 and 0.0020 respectively. This implies that the issuance of debt engages the firm to pay cash as interest and principal with availability of free cash flow and internally generated funds.

Liquidity is negatively related with investments and is not statistically significant and the regression co-efficient value is 0.01667. It implies that the failure of a firm to meet its obligation due to lack of sufficient liquidity will result in poor credit

worthiness loss of creditors confidence and this is not the case as shown by the results from the above table.

From the table it is observed that Tobin'Q is negatively related with investments and not statistically significant.

RESULT OF MEDIUM FIRMS'

Table No 2 Reveals that the regression results of medium firms. The calculated f value is greater than table value. Hence the selected variables are significantly associated with investment during the period. Further it shows that the leverage has no impact on investment in medium firm but it has negative relationship with investment during the period of study. In order to identify which methodology-pooling random effect or fixed effect regression model is most suitable, we perform two statistical tests, the first the LM test of the random effect model. The null hypothesis is that individual effect u_i is 0. The chi-square value is 4.15. Thus the null hypothesis is rejected at 1% level of significance. The results suggest that the rho effect is zero and the pooling regression is suitable in this case. The regression the co-efficient of leverage on medium firms from the pooling regression equal to 1.6543 and is not significant. The regression co-efficient on leverage from random and fixed effect model-0.7797 and-1.6543 respectively. The regression co-efficient from the pooling regression are greater than the those estimated from the random and fixed effect model suggesting that the individual effect of a firm leads to an estimation of the impact of leverage on investment.

Table 2:-Regression result of medium firms

Variables	Pooling	Fixed effect	Random effect	Fe with AR
Constant	2.7374 (7.94)	2.5853 (6.53)	2.7374 (7.94)	3.1669 (7.36)
leverage	-1.6543 (-0.56)	-0.7797 (-0.23)	-1.6543 (-0.56)	3.1893 (1.13)
Sales	-0.0016 (-1.85)	-0.0015 (-1.62)	-0.0016 (-1.85)	-0.0019 (-1.75)
ROA	-0.0021 (-0.59)	-0.0012 (-0.30)	-0.0021 (-0.59)	-0.0020 (-0.46)
Cash Flow	2.8863 (2.51)	2.3959 (1.82)	2.8863 (2.51)	1.0150 (0.95)
Liquidity	0.0907 (0.24)	-0.1078 (-0.23)	0.0907 (0.24)	-0.1226 (-0.38)
Retained earnings	0.0067 (5.94)	0.0078 (4.59)	0.0067 (5.94)	0.0048 (3.70)
Tobin's Q	-0.4580 (-1.59)	-0.4896 (-1.63)	-0.4580 (-1.59)	-0.3389 (-1.21)
LM Test	Chi ² (1)=4.15			
Hasuman Test	Chi ² (7)=55.74			
Observations (Groups)	150 -	150 10	150 10	150 10
Adjusted R ²	0.2465			

We conduct the Hausman specification test to compare the fixed effect and random effect models. If the model is correctly specified and if individual effects are uncorrelated with independent variable, the fixed effect and random effect estimates should not be statistically different. Further these

statistics are reported that the fixed effect model is most appropriate in estimating the investment equation because the R² value of fixed effect model is greater than random effect model.

Leverage is not statistically significant at 1% and five per cent level of significance and is negatively related with investment. This implies that leverage has no impact in medium firm's investment decision. It is because of inadequate cash flow and ploughing back of funds. Hence medium sized firms are making investment decision based on the internal financial resources. The table further reveals that the medium firms are under utilizing their fixed assets and it would affect the ability in generating the volume of sales and coefficient value is -0.0016 and is not statistically significant. The coefficient value of ROA is -0.0012 and is not statistically significant but negatively related with investment. The cash flow and Retained associated with in order earnings are positively associated with investment and they are statistically significant at 1% and 5% level of significance with investment. It indicates that higher the cash flow and retained funds higher will be the investment. Liquidity and Tobin's q are not statistically significant with investment the Tobin's Q also requested firm value and hence may be affected by leverage. But proxies in this do not influence the investment because the leverage has no impact on investment in medium firms.

RESULT OF LARGE FIRMS'

Table No 3 Shows that the regression results of large firms. The calculated f value is greater than table value. Hence the selected variables significantly associated with investments during the period of study. Further it shows that the leverage has no impact on investment in large firms but it has positive relationship with investments during the period of study. In order to identify which methodology-pooling, random effect fixed effect

regression model is most suitable. we perform two statistical test, the first the LM test of the random effect model. The null hypothesis is that individual effect u_i is 0. The chi-square value is 2.26. Thus null hypothesis is rejected @ 1% level of significance. The results suggest that the rho effect is not zero and the pooling regression is not suitable in this case. The regression coefficient of leverage on large firms from the pooling regression is equal to 23.7516 and is not significant. The regression coefficient on leverage from random effect and fixed effect model 9.5758 and 23.7516 respectively. The regression coefficient from the pooling regression Model is greater than those estimated from the random and fixed effect model suggesting that the individual effect of a firm leads to an estimation of the impact of leverage on investment.

We conduct the Hausman specification test to compare the fixed effect and random effect model if the model is correctly specified and if individual effect are correlated with independent variable the fixed and random effect are uncorrelated with independent variable the, fixed and random effect estimate should not be statistically different further these model is most appropriate that the fixed effect model is most appropriate in estimating the investment equation because the R^2 value of fixed effect model is greater than the random effect model.

The table also reveals that the coefficient value of variables like sales, ROA and Tobin's Q are negatively related with investment and also they are not significant in the leverage firms.

Cash flow and Retained earnings are positively associated with investment in large firms and are statistically significant it is because of heavy

demand for its product in national and international market. Liquidity is negatively related with investment and is not statistically significant with

investment. We conclude that the leverage is not influenced the investment decisions in large sized pharmaceutical firms in India.

Table 3:-Regression result of large firms

Variables	Pooling	Fixed effect	Random effect	Fe with AR
Constant	3.5155 (5.39)	2.7103 (3.81)	3.4388 (5.29)	3.5082 (5.63)
leverage	23.9805 (1.90)	9.5158 (0.67)	23.7516 (1.93)	21.9527 (1.70)
Sales	0.0023 (1.13)	0.0054 (2.40)	0.0025 (1.27)	0.0025 (1.26)
ROA	-0.0176 (-4.24)	-0.0092 (2.48)	-0.0160 (-4.01)	-0.0180 (-4.25)
Cash Flow	11.1396 (4.51)	16.1128 (7.57)	12.6212 (5.32)	10.3483 (4.05)
Liquidity	-4.9520 (-2.92)	-3.7888 (-1.49)	-5.4565 (-3.07)	-4.5578 (-2.62)
Retained earnings	-0.0005 (-0.27)	-0.0006 (-0.40)	-0.0002 (-0.14)	0.0003 (0.17)
Tobin's Q	0.5127 (0.83)	0.4301 (0.79)	0.5172 (0.86)	0.3846 (0.62)
LM Test	Chi ² (1)=2.26			
Hasuman Test	Chi ² (7)=59.44			
Observations (Groups)	40 -	40 10	40 10	40 10
Adjusted R ²	0.5474			

CONCLUSION:

This paper extends earlier empirical studies on the relationship between leverage and investment

in several dimensions. It verified the relationship for Top 25 Indian pharmaceutical firms that are quoted on the stock exchange of Mumbai for the year 1978 -

2009. Prior theoretical work posits that financial leverage can have either a positive or a negative impact on the value of the firm because of its influence on corporate investment decisions. The investigation is motivated by the theoretical work of Myers(1977) Jen Seen (1986), Stulz (1988,1990) and by an analytical work of McConnell and Servaes (1990). We examined whether financing consideration affects firm investment decisions. We found that leverage is positively related to the level of investment and that this positive effect is significantly stronger for firms with small firms and negative impact on medium firms but positive impact on large firms and this is not statistically significant. Further we inferred that the Indian pharmaceutical industry has heavy market demand for its product, so that Industry had enormous cash flow and bloughing

hack of funds. Hence we conclude that the leverage has no impact of pharmaceutical industry in India. Cash flow and Retained earning play significant role in determining the investment the decisions due to the change in the monetary policy of the country. Cash flow effect investment decisions due to the imperfections of the capital market and due to the fact internal financing is cheaper than external financing. These financing sources are far more important for small and highly leveraged firms. Our results support Hite (1977) Who found that leverage and investment are positively associated with given the level of financing if an investment increase would lower financial risk and hence the cost of bond financing.

REFERENCES

1. Aivazian, V.A Callen, J.L., 1980, "Corporate leverage and growth: the game theoretic issues", Journal of financial Economics 8, 379...399.
2. Beush, T., Pagan, A., 1980, "The language multiplier test and its applications to model specifications in econometrics", Review of economic studies 47, 239, 253. Economics of information and Uncertainty, University of Chicago press, Chicago, pp 107-140.
3. Cantor; Richard, (1990), "Effects of leverage on corporate investment and hiring decisions", Federal Bank of New York Quarterly Review, pp. 31-41.
4. Hausman, J.A., 1978, "Specification tests in econometric", Econometrica 46, 1251 – 1271.
5. Himmelberg, C.P., Hubbard, R.G., Palia, D., 1999, "Understanding the determinants of managerial ownership and the link between ownership and performance", Journal of financial economics, 53, 353-384.
6. Jensen; Michael, C., (1986), "Agency costs of free cash flow, corporate finance and takeovers", American Economic Review vol76, pp. 323-329.
7. Jensen, M.C., 1986, "Agency cost of free cash flow, corporate finance, and take-overs", American economic review 79, 323-329.
8. Johnson, Shane, A., (2003), "Debt maturity and the effects of growth opportunities and liquidity risk on leverage", Review of Financial Studies vol16, pp.209-236.
9. Kopcke and Howrey, (1994), "A panel study of investment: Sales, cash flow, the cost of capital, and leverage", New England Review, Jan/Feb., pp. 9-30
10. Lang, L.E.; Ofek, E.; Stulz, R., (1996), "Leverage, investment, and firm growth", Journal of Financial Economics, Vol40, pp. 3-29.
11. McConnell, John, J. and Servaes, H., (1995), "Equity ownership and the two faces of debt", Journal of Financial Economics, vol39, pp.131-157. .
12. Modigliani; Franco and Merton, H.; Miller (1958), "The cost of Capital, corporation finance, and the theory of investment", American Economic Review vol48, pp. 261-297.
13. Modigliani; Franco and Merton, H.; Miller (1963), "Corporate income taxes and the cost of capital", a correction, American Economic Review vol48, pp. 261-297.
14. Modigliani, F., Miller. M.H., 1958, "The cost of capital, corporation finance, and the theory of investment", American Economic Review 53, 433-443.

15. Myers , S., 1977, “*Determinants of corporate borrowing*”,
Journal of financial Economics,5, 147-175.

16. Whited, T. (1992). “*Debt, Liquidity constraints and corporate investment: Evidence from panel data*”, Journal of Finance vol 47, pp.1425-1461.

- Franklin John. S, The Principal, Nehru College of Management, Coimbatore, Tamilnadu, India.
Email : franklinjohn@rediffmail.com
- Muthusamy.K, Research Scholar, School of Management, Karunya University, Coimbatore, Tamilnadu, India.
Email: muthusamyphdmt@gmail.com

Effect of Nanofluid Concentration on the Performance of Circular Heat Pipe

M. G. Mousa

Abstract— The goal of this paper is to experimentally study the behavior of nanofluid to improve the performance of a circular heat pipe. Pure water and Al_2O_3 -water based nanofluid are used as working fluids. An experimental setup is designed and constructed to study the heat pipe performance under different operating conditions. The effect of filling ratio, volume fraction of nano-particle in the base fluid, and heat input rate on the thermal resistance is investigated. Total thermal resistance of the heat pipe for pure water and Al_2O_3 -water based nanofluid is also predicted. An experimental correlation is obtained to predict the influence of Prandtl number and dimensionless heat transfer rate, K_q on thermal resistance. Thermal resistance decreases with increasing Al_2O_3 -water based nanofluid compared to that of pure water. The experimental data is compared to the available data from previous work. The agreement is found to be fairly good.

Index Terms— Heat Pipe, Thermal Performance, Nanofluids

Nomenclature

A	Surface area, m^2
FR	Filling ratio
I	Electric current, Amp
k	thermal conductivity, $W/m.K$
N	Number of thermocouples
Q	Input heat rate, W
q	Heat flux, W/m^2
R	Total thermal resistance of heat pipe, K/W
T	Temperature, K
V	Applied voltage, Volt
X	Horizontal coordinate parallel to the test section, mm

Greek Symbols

ϕ	Volume fraction of nanoparticles, %
ρ	Density, kg/m^3
μ	Dynamic viscosity, $N.s/m^2$

Subscript

c	Condenser
e	Evaporator
ef	Effective
e	Evaporator
l	Liquid
m	Base fluid
n	Nanofluid
p	Particles
water	Pure water

Dimensionless Numbers

RR Reduction factor in thermal resistance

K_q Dimensionless heat transfer rate, $\frac{K_{ef} \times L_e \times \Delta T}{Q}$

Pr Prandtl number, $\frac{\mu_{ef} \times Cp_{ef}}{K_{ef}}$

INTRODUCTION

To solve the growing problem of heat generation by electronic equipment, two-phase change devices such as heat pipe and thermosyphon cooling systems are now used in electronic industry. Heat pipes are passive devices that transport heat from a heat source to a heat sink over relatively long distances via the latent heat of vaporization of a working fluid. The heat pipe generally consists of three sections; evaporator, adiabatic section and condenser. In the evaporator, the working fluid evaporates as it absorbs an amount of heat equivalent to the latent heat of vaporization. The working fluid vapor condenses in the condenser and then, returns back to the evaporator. Nanofluids, produced by suspending nano-particles with average sizes below 100 nm in traditional heat transfer fluids such as water and ethylene glycol provide new working fluids that can be used in heat pipes. A very small amount of guest nano-particles, when uniformly and suspended stably in host fluids, can provide dramatic improvement in working fluid thermal properties. The goal of using nanofluids is to achieve the highest possible thermal properties using the smallest possible volume fraction of the nano-particles (preferably < 1% and with particle size < 50 nm) in the host fluid.

Kaya et al. [1] developed a numerical model to simulate the transient performance characteristics of a loop heat pipe.

Kang et al. [2], investigated experimentally, the performance of a conventional circular heat pipe provided with deep grooves using nanofluid. The nanofluid used in their study was aqueous solution of 35 nm diameter silver nano-particles. It is reported that, the thermal resistance decreased by 10-80% compared to that of pure water.

Pastukhov et al. [3], experimentally, investigated the performance of a loop heat pipe in which the heat sink was an external air-cooled radiator. The study showed that the use of additional active cooling in combination with loop heat pipe increases the value of dissipated heat up to 180 W and decreases the system thermal resistance down to 0.29 K/W.

Chang et al. [4] investigated, experimentally, the thermal performance of a heat pipe cooling system with thermal resistance model. An experimental investigation of thermosyphon thermal performance considering water and dielectric heat transfer liquids as the working fluids was performed by **Jouhara et al. [5]**. The copper thermosyphon was 200 mm long with an inner diameter of 6 mm. Each thermosyphon was charged with 1.8 ml of working fluid and tested with an evaporator length of 40 mm and a condenser length of 60 mm. The thermal performance of the water charged thermosyphon is compared with the three other working fluids (FC-84, FC-77 and FC-3283). The parameters considered were the effective thermal resistance as well as the maximum heat transport. These fluids have the advantage of being dielectric which may be better suited for sensitive electronics cooling applications. Furthermore, they provide adequate thermal performance up to approximately 50 W, after which liquid entrainment compromises the thermosyphon performance.

Lips et al. [6], studied experimentally, the performance of

plate heat pipe (FPHP). Temperature fields in the heat pipe were measured for different filling ratios, heat fluxes and vapor space thicknesses. Experimental results showed that the liquid distribution in the FPHP and consequently its thermal performance depends strongly on both the filling ratio and the vapor space thickness. A small vapor space thickness induces liquid retention and thus reduces the thermal resistance of the system. Nevertheless, the vapor space thickness influences the level of the meniscus curvature radii in the grooves and hence reduces the maximum capillary pressure. Thus, it must be, carefully, optimized to improve the performance of the FPHP. In all the cases, the optimum filling ratio obtained, was in the range of one to two times the total volume of the grooves. A theoretical approach, in non-working conditions, was developed to model the distribution of the liquid inside the FPHP as a function of the filling ratio and the vapor space thickness.

Das et al. [7-8] and **Lee et al. [9]** found great enhancement of thermal conductivity (5-60%) over the volume fraction range of 0.1 to 5%.

All these features indicate the potential of nanofluids in applications involving heat removal. Issues, concerning stability of nanofluids, have to be addressed before they can be put to use. Ironically, nanofluids of oxide particles are more stable but less effective in enhancing thermal conductivity in comparison with nanofluids of metal particles.

The aim of the present work is to investigate, experimentally, the thermal performance of a heat pipe. The affecting parameters on thermal performance of heat pipe are studied. The type of working fluid (pure water and Al₂O₃-water based nanofluid), filling ratio of the working fluid, volume fraction of nano-particles in the base fluid, and heat input rate are considered as experimental parameters. Empirical correlation for heat pipe thermal performance, taking into account the various operating parameters, is presented.

2. Experimental Setup and Procedure

A schematic layout of the experimental test rig is shown in Fig.1. This research adopts pure water and Al₂O₃-water based nanofluid as working fluids. The size of nano-particles is 40 nm. The test nanofluid is obtained by dispersing the nano-particles in pure water. The working fluid is charged through the charging line (6). In the heat pipe, heat is generated using an electric heater (12). The vapor generated in the evaporator section (8) is moved towards the condenser section (4) via an adiabatic tube (5) whose diameter and length are 20 mm and 40 mm, respectively. Both evaporator and condenser sections have the size of 40 mm-diameter and 60 mm-height. The condensate is allowed to return back to evaporator section by capillary action "wick structure" through the adiabatic tube. The surfaces of the evaporator section, adiabatic section, and condenser section sides are covered with 25 mm-thick glass wool insulation (3). Seventeen calibrated cooper-constantan thermocouples (T-type) are glued to the heat pipe surface and distributed along its length to measure the local temperatures (Fig. 2). Two thermocouples are used to

measure ambient temperature. All thermocouples are connected to a digital temperature recorder via a multi-point switch. The non condensable gases are evacuated by a vacuum pump. The heat pipe is evacuated to 0.01 bar via the vacuum line (10). The power supplied to the electric heater (12) is measured by a multi-meter (13). The input voltage was adjusted, using an autotransformer (2). The voltage drops across the heater were varied from 5 to 45 Volts. The A.C. voltage stabilizer (1) is used to ensure that there is no voltage fluctuation during experiments. The pressure inside the evaporator was measured by a pressure gage with a resolution of 0.01 bar.

Thermocouples (with the uncertainty lower than 0.20 °C) are distributed along the surfaces of the heat pipe sections as follows: six thermocouples are attached to the evaporator section, two thermocouples are attached to the adiabatic section, and nine thermocouples are attached to the condenser section. The obtained data for temperatures and input heat rate are used to calculate the thermal resistance.

One can define the filling ratio, FR, as the volume of charged fluid to the total evaporator volume. The working fluid is charged at 30 °C.

The effects of working fluid type, filling ratio, volume fraction of nano-particles in the base fluid, and heat input rate on the thermal performance of the heat pipe are investigated in the experimental work. The experimental runs are executed according to the following steps:

1. The heat pipe is evacuated and charged with a certain amount of working fluid
2. The supplied electrical power is adjusted manually at the desired rate using the autotransformer.
3. The steady state condition is achieved after, approximately, one hour of running time using necessary adjustments to the input heat rate. After reaching the steady state condition, the readings of thermocouples are recorded, sequentially, using the selector switch. The voltage of the heater is measured to determine the value of applied heat flux. Finally, the pressure inside the evaporator is recorded.
4. At the end of each run, power is changed and step 3 is repeated.
5. Steps 1 through 4 are repeated using with another adjusted amount of working fluid. The filling ratios, FR, used are 0.2, 0.4, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.80 and 1.0.

Pure water and Al₂O₃-water based nanofluid are used as working fluids. Steps 1 through 5 are repeated for Al₂O₃-water based nanofluid using several values of volume fractions of nano-particles. The volume fractions used are 0.25%, 0.4%, 0.5%, 0.6%, 0.75%, 1.0% and 1.5%, respectively.

3. Data Reduction

Although heat pipes are very efficient heat transfer devices, they are subject to a number of heat transfer limitation. For high heat flux heat pipes operating in low to moderate temperature range, the capillary effect and boiling limits are commonly the dominant factor. For a given capillary wick structure and working fluid combination, the pumping ability of the capillary structure to provide the circulation for a given working fluid is limited. In order to

maintain the continuity of the interfacial evaporation, capillary pressure must satisfy the following relation [2]

$$\Delta P \geq \Delta P_l + \Delta P_v + P_e + P_c \quad (1)$$

Boiling limit is directly related to bubble formation in the liquid. In order that a bubble can exist and grow in liquid, a certain amount of superheat is required. Accurately characterizing the thermal power transfer, Q is a complicated task because it is difficult to accurately quantify the energy loss to the ambient surroundings. Therefore, the whole surface of the heat pipe is well insulated so that the rate of heat loss can be ignored. The heat input rate can be calculated using the supplied voltage and measured current such that,

$$Q = I \times V \quad (2)$$

Where V and I are the applied voltage in Volt and current in Amp, respectively.

The experimental determination of the thermal performance of the heat pipe requires accurate measurements of evaporator and condenser surface temperatures as well as the power transferred. Calculating evaporator and condenser temperatures is, relatively, a straightforward task. They are obtained by simply averaging the temperature measurements along the evaporator and condenser surfaces. Thus, evaporator and condenser temperatures can be expressed as:

$$\bar{T}_e = \frac{\sum_{i=1}^{i=N_e} T_{ei}}{N_e}, \quad \bar{T}_c = \frac{\sum_{i=1}^{i=N_c} T_{ci}}{N_c} \quad (3)$$

Where; N_e, N_c are the number of thermocouples on the evaporator and condenser, respectively. The obtained data for temperatures and heat input rate are then used to calculate the thermal resistance using the following relation,

$$R = \frac{(\bar{T}_e - \bar{T}_c)}{Q} \quad (4)$$

One can assume that the nano-particles are well dispersed within the base-fluid, so the effective physical properties are described by classical formulas which are mentioned by Das et al. [10] as;

$$\rho_{ef} = (1 - \phi)\rho_m + \phi\rho_p$$

The effective dynamic viscosity of nanofluids can be calculated using different existing equations that have been obtained for two-phase mixtures. The following relation is the well-known Einstein's equation for a viscous fluid containing a dilute suspension of small, rigid, spherical particles.

$$\mu_{ef} = \mu_m (1 + 2.5 \phi) \quad (6)$$

$$k_{ef} = k_m (1 + 7.4 \phi)$$

Where, ϕ is the ratio of the volume of the nano-particle to the volume of the base fluid.

The symbols K_m and ϕ are the base fluid thermal conductivity and volume fraction of the nano-particle in the base fluid respectively.

The relevant thermo-physical properties of the solid nano-particles (Al_2O_3) used in the present study are $C_{p_p} = 773$ J/kg. $^{\circ}C$, $\rho = 3880$ kg/ m^3 , and $k_p = 36$ W/ $m.^{\circ}C$, which are mentioned in previous work [10].

One can calculate the reduction factor in total thermal resistance of heat pipe charged with nanofluid by referring its thermal resistance to that charged with pure water, expressed as;

$$RR = (R_{water} - R_n) / R_{water} \quad (7)$$

The maximum relative error in total thermal resistance is about + 11.9 %.

4. Result and Discussion

Experiments are performed on heat pipe considering two different working fluids; pure water and Al_2O_3 -water based nanofluid. In both cases, the effect of heat input rate, Q , and filling ratio, FR on its performance are investigated. Moreover, in case of the nanofluid, the effect of varying volume fraction of the nano-particles in the base fluid, ϕ , on the thermal performance of this heat pipe is also predicted. The values of local surface temperatures along all sections of the heat pipe are measured ($0 \leq X \leq 120$ mm), where X is measured from the beginning of the evaporator section. The range $0.0 \leq X \leq 40$ mm represents the evaporator section, $40 < X \leq 80$ mm represents the adiabatic section and $80 < X \leq 120$ mm represents the condenser section.

Figure 3 and 4 illustrate the surface temperature along the heat pipe for two values of heat input rate (40 and 60 W) and same filling ratio, FR of 0.45. As expected, the surface temperature decreases with increasing the

distance from the evaporator section due to the heat exchange between the heat pipe surface and working fluid. It is clear that the surface temperature decreases with increase of nano-particle concentration, ϕ . As expected, the surface temperature decreases with increasing the distance from the evaporator along heat pipe.

Figure 5 illustrates the variation of the total thermal resistance of the heat pipe, R with the filling ratio for Al_2O_3 -water based nanofluid ($\phi = 0, 0.5\%$ and 1.2%) at heat input rate of 40 W. As shown in the figure, R decreases with the increase of the filling ratio up to a value of FR equals to 0.40, after which R starts to increase with the increase of FR due to increasing liquid inside evaporator. It can be also noticed that the thermal resistance, R is inversely proportional to concentration of nano-particles, ϕ .

Figure 6 shows the variation of the total thermal resistance of the heat pipe, R with volume fraction of nano-particle in the base fluid, ϕ at two different heat input rates, Q (40 and 60 W). Over the tested range of ϕ , while keeping the filling ratio of 0.45, the percentage enhancement in R reaches up to 62.60% at $\phi = 1.2\%$ compared to its value when

using pure water.

The effect of adding nano-particles on the thermal performance of the heat pipe is more evident if the data is expressed as a plot of the reduction rate in total thermal resistance, RR versus ϕ , as shown in figure 7, i.e. the enhancement of thermal performance is increasing with the increase of nano-particles concentration, ϕ . The addition of nano-particles has illustrated that during nucleate boiling some nano-particles deposit on the heated surface to form a porous layer. This layer improves the wet ability of the surface considerably. The thermal conductivity of the working fluid is also preferably high in order to minimize the temperature gradient.

The obtained heat transfer data is correlated as the following relation:

$$R = 0.294 [K_q^{-0.596} FR^{1.273} Pr^{-0.0532}] \quad (8)$$

The error in calculated thermal resistance is predicted by the above suggested correlation is around $\pm 5\%$, as shown in Fig. 8.

4.1. Comparison with the available literature

Figure 9 shows a comparison between the present experimental results with those reported by Kang [2] in case of using pure water as a working fluid with FR equals to 0.5. It can be observed that the present experimental results for the used two working fluids have the same trend as those reported by Kang [2]. The difference between both results when using pure water may be attributed to the difference in heat pipe geometry and uncertainty in measurements.

Figure 10 shows a comparison between the present experimental results of the total resistance, R using Al_2O_3 -water based nanofluid of $\phi = 0.5\%$ to those reported by Lips and Lefevre [6] who used a heat pipe charged with n-pentane nanofluid of $\phi = 0.5\%$. One can see that the thermal resistance of the heat pipe decreases with increasing filling ratio up to a value of $FR = 0.45$, where it starts to increase with increasing the filling ratio. It can be also observed that the present experimental results are, a little bit, higher than those reported by Lips et al [6], but they have the same trend. The discrepancies in both results may be due to the differences in dimensions of the tested heat pipes as well as due to the type of working fluids used.

5. Conclusions

Using heat pipes and based on the nanofluid literature, particularly those related to the optimum operating condition, the thermal performance enhancement of heat pipes charged with the nanofluids indicates the potential of the nanofluid use as substitute of conventional fluids. This finding makes the nanofluid more attractive as a cooling fluid for devices with high power intensity. A compact heat pipe is thermally tested with two different working fluids; pure water and Al_2O_3 -water based the nanofluid. The thermal performance of this heat pipe is predicted under different operating conditions including heat input rate, filling ratio, and volume fraction of the nano-particle in water. From the obtained data and its discussion, the following conclusions may be drawn:

1- The optimum filling ratio of charged fluid in the tested heat pipe was about 0.45 to 0.50 for both pure water and

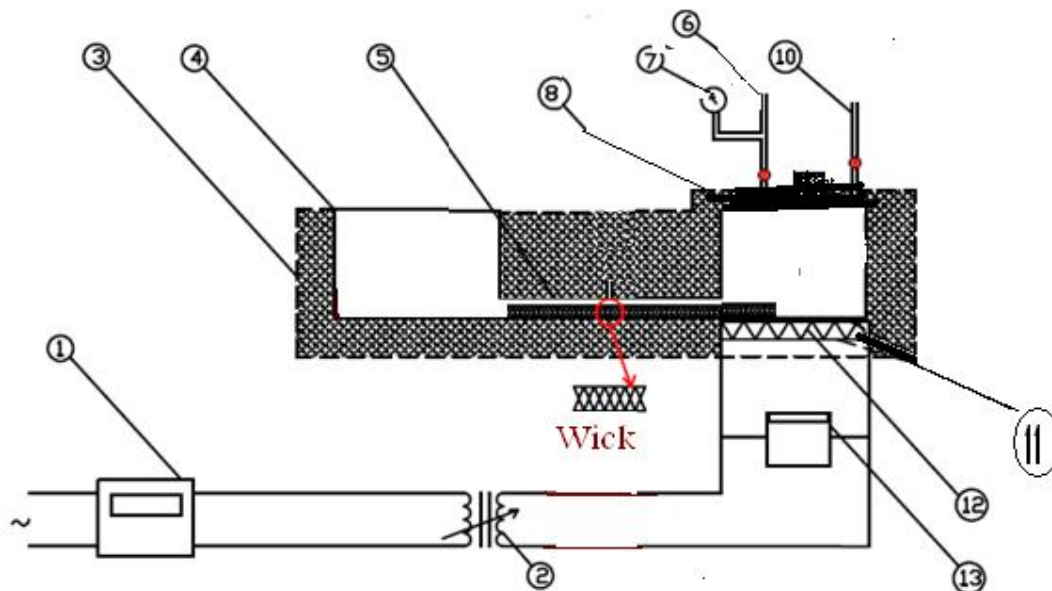
Al₂O₃-water based the nanofluid, respectively.

2- By increasing concentration of the nanofluid, the thermal performance of heat pipe can be decreased

It can be said that better ability to manage thermal properties of working fluid translates into greater energy transport, smaller and lighter thermal systems. This may be applied to cooling of super computers.

References

- [1] **T. Kaya, R. Pe´rez, C. Gregori and A.Torres**, "Numerical simulation of transient operation of loop heat pipes", Applied Thermal Engineering, vol.28, 2008, pp.967-974.
- [2] **S. W. Kang, W. C. Wei, S. H. Tsai, and S. Y. Yang**, "Experimental investigation of silver nanofluid on heat pipe thermal performance", Applied Thermal Engineering, vol.26, 2006, pp.2377–2382.
- [3] **V.G. Pastukhov, Y. F. Maidanik, C.V. Vershinin, and M.A. Korukov**, "Miniature loop heat pipes for electronics cooling", Applied Thermal Engineering, vol.23, 2000, pp.1125-1135.
- [4] **Y. W. Chang, C. H. Cheng, J. C. Wang, and S. L. Chen**, "Heat pipe for cooling of electronic equipment" Energy Conversion and Management, 2008.
- [5] **H. Jouhara¹, O. Martinet, and A.J. Robinson**, "Experimental Study of Small Diameter Thermosyphons Charged with Water", FC-84, FC-77 & FC-3283- 5th European Thermal-Sciences Conference, The Netherlands, 2008.
- [6] **S. Lips, F. Lefèvre, and J. Bonjou**, "Combined effects of the filling ratio and the vapour space thickness on the performance of a flat plate heat pipe", International Journal of Heat and Mass Transfer vol. 53, 2010, pp. 694–702.
- [7] **S. K. Das, U.S. Choi, W. Yu, and T. Pradeep**, "Nanofluid Science and Technology", Wiley-Interscience, 2007.
- [8] **Das, S. K., Putra, N., Thiesen, P., and Roetzel, W.**, "Temperature dependence of thermal conductivity enhancement for nanofluids", J. Heat Transfer vol.125, 2003, pp. 567-574.
- [9] **R.B. Mansour, N. Galanis, and C.T. Nguyen**, " Effect of uncertainties in physical properties on forced convection heat transfer with nanofluids," Applied Thermal Engineering Vol. 27, 2007, pp. 240–249.
- [10] **S. Lee, S.U.S Choi, S. Li., and J.A. Eastman**, "Measuring thermal conductivity of fluids containing oxide nanoparticles," J. Heat Transfer vol. 121, 1999, pp. 280-289.



1. Stabilizer	2. Autotransformer	3. Insulation	4. condenser	5. Adiabatic tube
6. Charge line	7. Pressure gage	8. Evaporator	10. Vacuum line	
11. Mica sheet	12. Electric heater	13. Multi-meter		

Figure 1 Schematic layout of the test rig

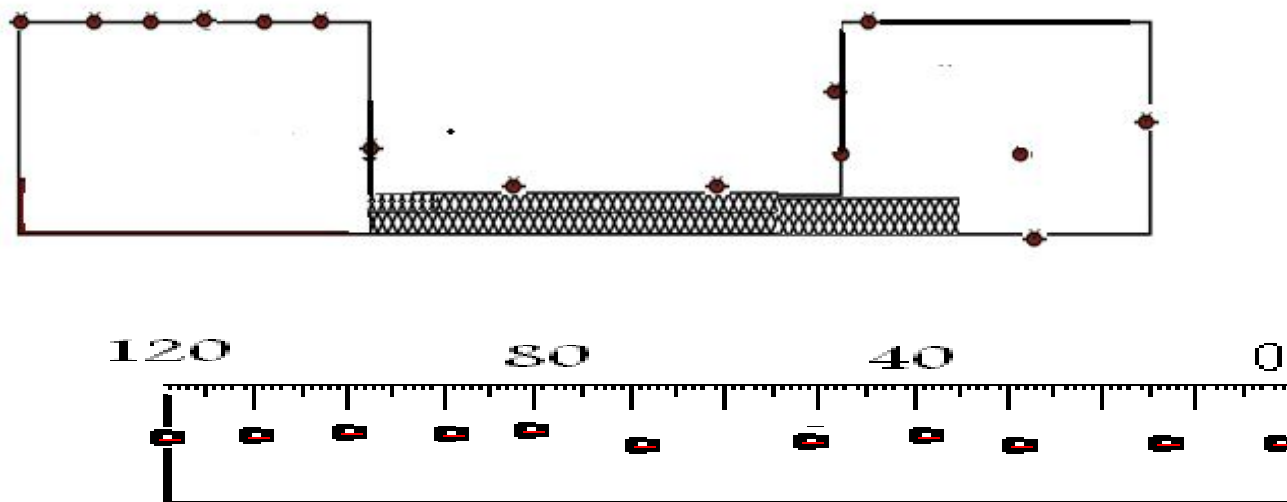


Figure 2 Thermocouples distribution along the heat pipe sections

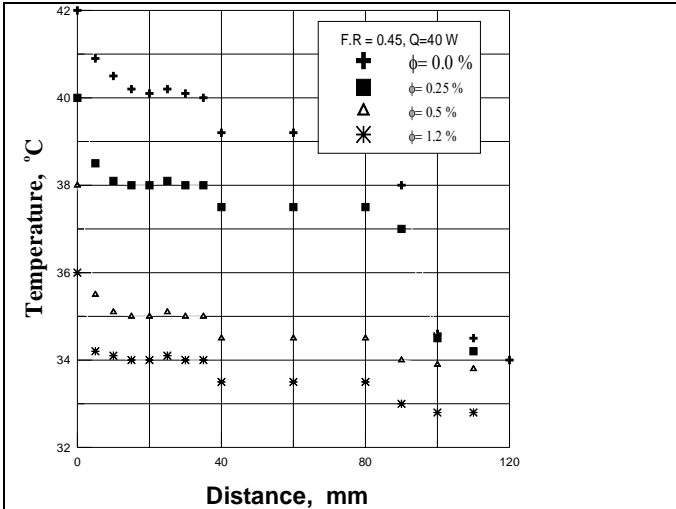


Figure 3 Temperature distribution along the heat pipe surface for different nanofluid concentration

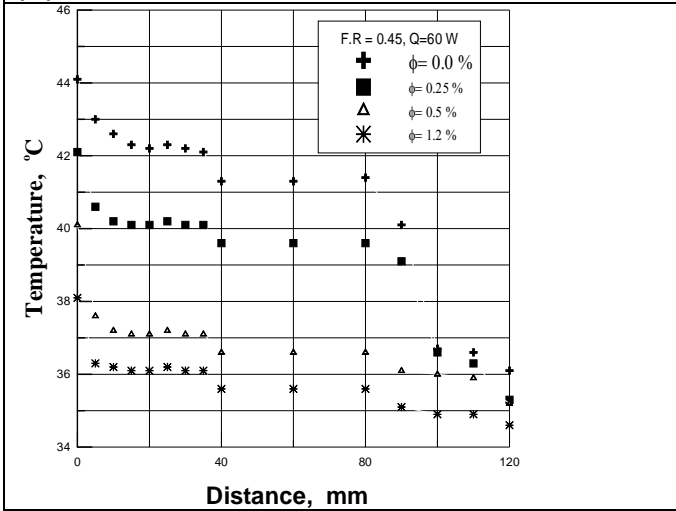


Figure 4 Temperature distribution along the heat pipe surface for different nanofluid concentration

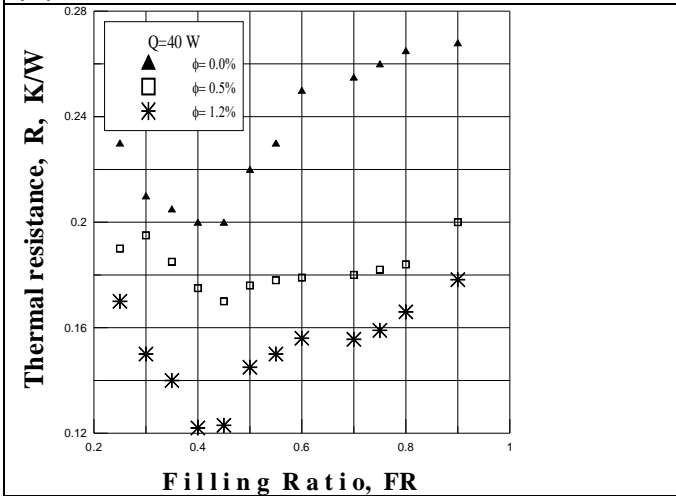


Figure 5 Variation of thermal resistance with filling ratio at different heat input rates

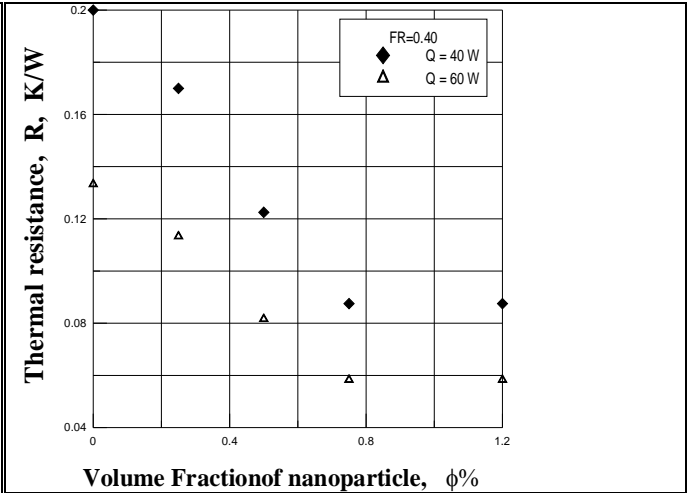


Figure 6 Variation of thermal resistance with volume fraction of nano-particles.

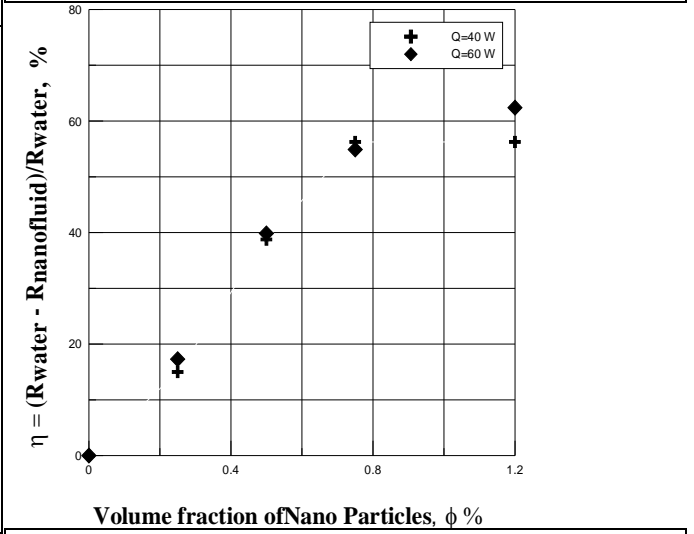


Figure 7 Reduction factor of total thermal resistance at different volume fraction of nano-particles, phi

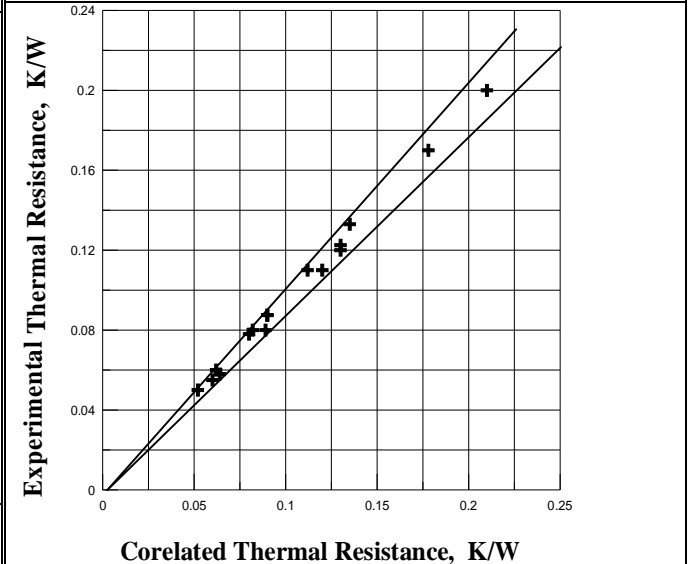


Figure 8 Experimental Nusselt number versus correlated

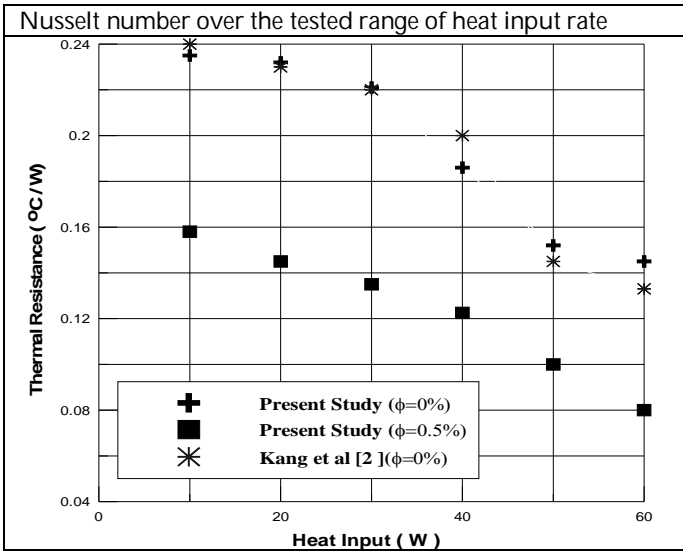


Figure 9 Comparison of the present results with available literature at different value of heat input rate

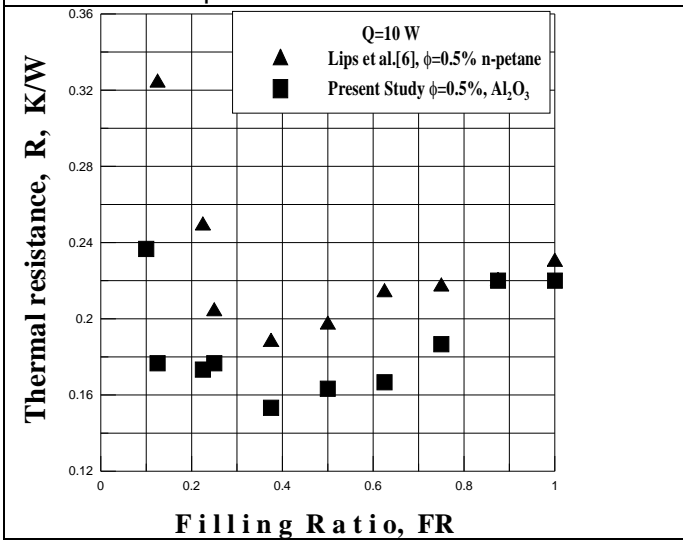


Figure 10 Comparison between present data and previous one for the variation of thermal resistance with filling ratio

Enhancement of Person Identification using Iris Pattern

Vanaja roselin.E.Chirchi, Dr.L.M.Waghmare, E.R.Chirchi

Abstract — The biometric person identification technique based on the pattern of the human iris is well suited to be applied to access control. Security systems having realized the value of biometrics for two basic purposes: to verify or identify users. In this busy world, identification should be fast and efficient. In this paper we focus on an efficient methodology for identification and verification for iris detection using Haar wavelet and the classifier used is Minimum hamming distance, even when the images have obstructions, visual noise and different levels of illuminations.

Index Terms—Biometrics, Iris identification, Haar wavelet, occluded images, veriEye.

1. Introduction

Biometrics which refers to identifying an individual by his or her physiological or behavioral characteristics has capability to distinguish between authorized user and an imposter. An advantage of using biometric authentication is that it cannot be lost or forgotten, as the person has to be physically present during at the point of identification process [9]. Biometrics is inherently more reliable and capable than traditional knowledge based and token based techniques. The commonly used biometric features include speech, fingerprint, face, Iris, voice, hand geometry, retinal identification, and body odor identification [10] as in Fig.1

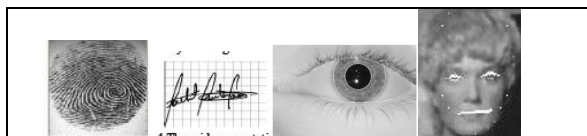


Fig. 1: Examples of Biometrics

To choose the right biometric to be highly fit for the particular situation, one has to navigate through some complex vendor products and keep an eye on future developments in technology and standards. Here comes a list of Biometrics with comparatives:

Facial Recognition: Facial recognition records the spatial geometry of distinguishing features of the face. Different vendors use different methods of facial recognition, however, all focus on measures of key features of the face. Facial recognition has been used in projects to identify card counters or other undesirables in casinos, shoplifters in stores, criminals and terrorists in urban areas. This biometric system can easily spoof by

the criminals or malicious intruders to fool recognition system or program. Iris cannot be spoofed easily.

Palm Print: Palm print verification is a slightly modified form of fingerprint technology. Palm print scanning uses an optical reader very similar to that used for fingerprint scanning; however, its size is much bigger, which is a limiting factor for use in workstations or mobile devices.

Signature Verification: It is an automated method of examining an individual's signature. This technology is dynamic such as speed, direction and pressure of writing, the time that the stylus is in and out of contact with the "paper". Signature verification templates are typically 50 to 300 bytes. Disadvantages include problems with long-term reliability, lack of accuracy and cost.

Fingerprint: A fingerprint as in Fig.1 recognition system constitutes of fingerprint acquiring device, minutia extractor and minutia matcher. As it is more common biometric recognition used in banking, military etc., but it has a maximum limitation that it can be spoofed easily. Other limitations are caused by particular usage factors such as wearing gloves, using cleaning fluids and general user difficulty in scanning.

Iris Scan: Iris as shown in Fig.2 is a biometric feature, found to be reliable and accurate for authentication process comparative to other biometric feature available today. As a result, the iris patterns in the left and right eyes are different, and so are the iris patterns of identical twins. Iris templates are typically around 256 bytes. Iris scanning can be used quickly for both identification and verification applications because of its large number of degrees of freedom. Iris as in Fig. 2 is like a diaphragm between the pupil and the sclera and

its function is to control the amount of light entering through the pupil. Iris is composed of elastic connective tissue such as trabecular meshwork. The agglomeration of pigment is formed during the first year of life, and pigmentation of the stroma occurs in the first few years [7][8].

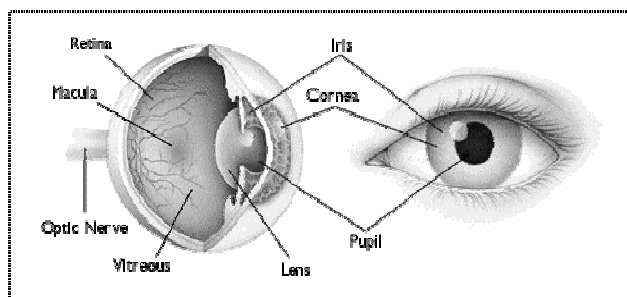


Fig. 2: Structure of Eye

The highly randomized appearance of the iris makes its use as a biometric well recognized. Its suitability as an exceptionally accurate biometric derives from [4]:

- i. The difficulty of forging and using as an imposter person;
- ii. It is intrinsic isolation and protection from the external environment;
- iii. It's extremely data-rich physical structure;
- iv. Its genetic properties—no two eyes are the same. The characteristic that is dependent on genetics is the pigmentation of the iris, which determines its color and determines the gross anatomy. Details of development, that are unique to each case, determine the detailed morphology;
- v. its stability over time; the impossibility of surgically modifying it without unacceptable risk to vision and its physiological response to light, which provides a natural test against artifice.

After the discovery of iris, John G. Daugman, a professor of Cambridge University [8], [9], suggested an image-processing algorithm that can encode the iris pattern into 256 bytes based on the Gabor transform.

In general, the iris recognition system is composed of the following five steps as depicted in Fig. 3 According to this flow chart, preprocessing including image enhancement.

2. Image Acquisition.

An image of the eye to be analyzed must be acquired first in digital form suitable for analysis. In

further implementation we will be using Chinese academy of science-Institute of automation (CASIA) iris image database available in the public domain [7].

3. Preprocessing

3.1 Locating Iris

The first processing step consists in locating the inner and outer boundaries of the iris and second step to normalize iris and third step to enhance the original image as in Fig.4 [4][6]. The Daugman's system, Integro differential operators as in (1) is used to detect the center and diameter of iris and pupil respectively.

$$\max(r, x_0, y_0) = \left\{ \frac{\partial}{\partial r} \int_0^{2\pi} I(r \cos \theta + x_0, r \sin \theta + y_0) \right\} \dots (1)$$

Where (x_0, y_0) denotes the potential center of the searched circular boundary, and r its radius.

3.2 Cartesian to polar reference transform

Cartesian to polar reference transform suggested by J.Daugman authorizes equivalent rectangular representation of the zone of interest as in Fig.4, remaps each pixel in the pair of polar co-ordinates (r, θ) where r and θ are on interval $[0,1]$ and $[0,\pi]$ respectively. The unwrapping is formulated as in (2) [1].

$$I(x(r, \theta), y(r, \theta)) \rightarrow I(r, \theta) \dots (2)$$

Such that

$$\left. \begin{aligned} x(r, \theta) &= (1-r)x_p(\theta) + rx_i(\theta), \\ y(r, \theta) &= (1-r)y_p(\theta) + ry_i(\theta) \end{aligned} \right\} \dots (3)$$

where $I(x, y)$, (x, y) , (r, θ) , (x_p, y_p) , (x_i, y_i) are the iris region, Cartesian coordinates, corresponding polar coordinates, coordinates of the pupil, and iris boundaries along the θ direction, respectively.

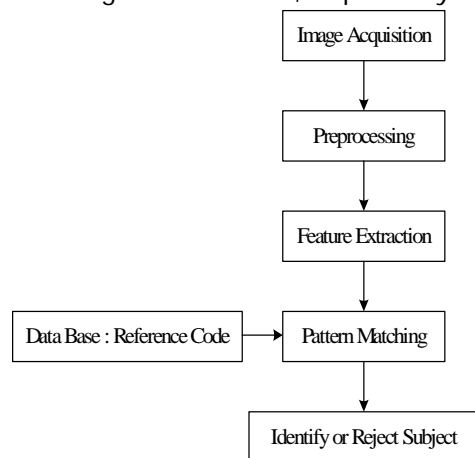
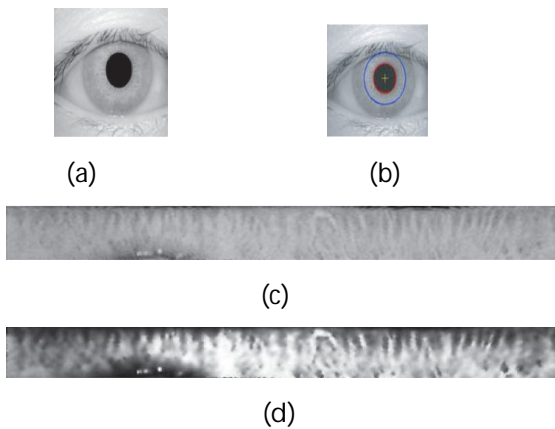


Fig. 3: General steps of the iris recognition system

Fig.4: (a) Original image; (b) localized iris; (c) normalized iris;
And (d) enhanced iris.

4. Feature Extraction

The most important step in automatic iris recognition is the ability of extracting some unique attributes from iris, which help to generate a specific code for each individual. Gabor and wavelet transforms are typically used for analyzing the human iris patterns and extracting features from them, In Fig. 5(a) [1],[2],[4] a conceptual chart of basic decomposition steps for an image is depicted. The approximation coefficients matrix cA and details coefficients matrices cH , cV and cD (horizontal, vertical, and diagonal, resp.) Obtained by wavelet decomposition of the input image are shown in Fig. 5(b) [4],[11],[12]. The definitions used in the chart are as follows.

- (i) $C \downarrow$ denote downsample columns—keep the even indexed columns.
- (ii) $D \downarrow$ denote downsample rows—keep the even-indexed rows.
- (iii) Lowpass D denotes the decomposition lowpass filter.
- (iv) Highpass D denotes the decomposition highpass filter.
- (v) The blocks under “Rows” convolve with filter of block the rows of entry.
- (vi) The blocks under “Columns” convolve with filter of block the columns of entry.
- (vii) I_i denotes the input image.

5. Pattern Matching

By comparing similarities between feature vectors of two irises to determine that they are accepted or rejected. Since feature vector is binary, the matching

process will be fast and simple accordingly. Performance of classifiers is based on minimum Hamming Distance (MHD) as in (4).

$$HD = \text{XOR}(\text{codeA}, \text{codeB}) \quad \text{--- (4)}$$

Where codeA and codeB are the templates of two images. When an iris image is captured in system, the designed classifier compares it with the whole images in each class. The Hamming distance (HDs) between input images and images in each class are calculated, then the two different classifiers are being applied as follows [1][4].

- I. In the first classifier, the minimum HD between input iris code and codes of each class is computed.
- II. In the second classifier, the harmonic mean of the n HDs that have been recorded yet is assigned to the class as in (5)[4].

$$HM = \frac{\text{length}(\text{code})}{\sum_{i=1}^n \text{length}(\text{code}) (1/\text{code}(i))} \quad \text{--- (5)}$$

6. Identification and Verification

Identification and verification modes are two main goals of every security system based on the needs of the environment. In the verification stage, the system checks if the user data that was entered is correct or not (e.g., username and password) but in the identification stage, the system tries to discover who the subject is without any input information. Hence, verification is a one-to-one search but identification is a one-to-many comparison.

7. Application

Implementations of image processing above five steps as in Fig.3 is used in veriEye iris recognition technology for healthcare patient identification system in 2008[5].Algorithm advanced iris segmentation, enrollment and matching which is as follows [5].

Robust eye iris detection: Irises are detected even when the images have obstructions, visual noise and different levels of illumination. Lighting reflections, eyelids and eyelashes obstructions are eliminated. Images with narrowed eyelids or eyes that are gazing away are also accepted using wavelet algorithm[5].

Automatic interlacing detection and correction: The correction results in maximum quality of iris features templates from moving iris images.

Gazing-away eyes: A gazing-away iris image is correctly detected, segmented and transformed as if it were looking directly into the camera.

Correct iris segmentation: is achieved under these conditions:

- i. *Perfect circles fail.* VeriEye uses active shape models that more precisely model the contours of the eye, as perfect circles do not model iris boundaries.
- ii. *The centers of the iris inner and outer boundaries are different* Fig. 8. The iris inner boundary and its center are marked in red; the iris outer boundary and its center are marked in green.
- iii. *Iris boundaries are definitely not circles and even not ellipses* Fig. 9. and especially in gazing-away iris images.
- iv. *Iris boundaries seem to be perfect circles.* The recognition quality can still be improved if boundaries are found more precisely Fig.10. Compared to perfect circular white contours
- v. *Fast matching.* Configurable matching speed varies from 50,000 to 150,000 comparisons per second. The highest speed still preserves nearly the same recognition quality Fig. 11.
- vi. *Reliability.* VeriEye 2.0 algorithm shows excellent performance when tested on all publicly available datasets

7.1 Technical Specification[5]

Minimal radius of circle containing full iris texture	64 pixels
Iris rotation tolerance	±15 degrees
Recommended iris image capture spectral region	Near-infrared
Iris template extraction time	0.5 sec
Matching speed	50,000 - 150,000 irises/sec

Size of one record in a database	2.3 Kbytes
Maximum database size	Unlimited

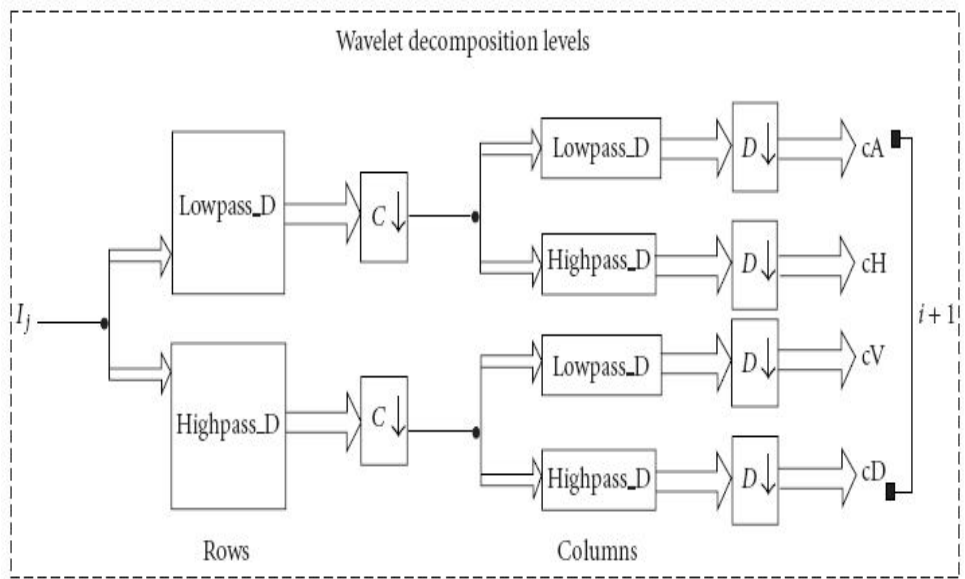
8. CONCLUSION

The research work is to enhance the algorithm for efficient person identification for other area of applications by increasing FRR more than 0.33% as the VeriEye algorithm results with FRR 0.32% and FAR 0.001%. Wavelets iris recognition algorithm is suitable for reliable, fast and secure person identification. Wavelet, Gabor filter and hamming distance are used in veriEye algorithm for robust and fast matching for healthcare application for patient identification. Research work will also focus on the algorithm for rapid and accurate iris identification even if the images are occlude further algorithm will also focus on robust iris recognition, even with gazing-away eyes or narrowed eyelids which solves all the security related problems.

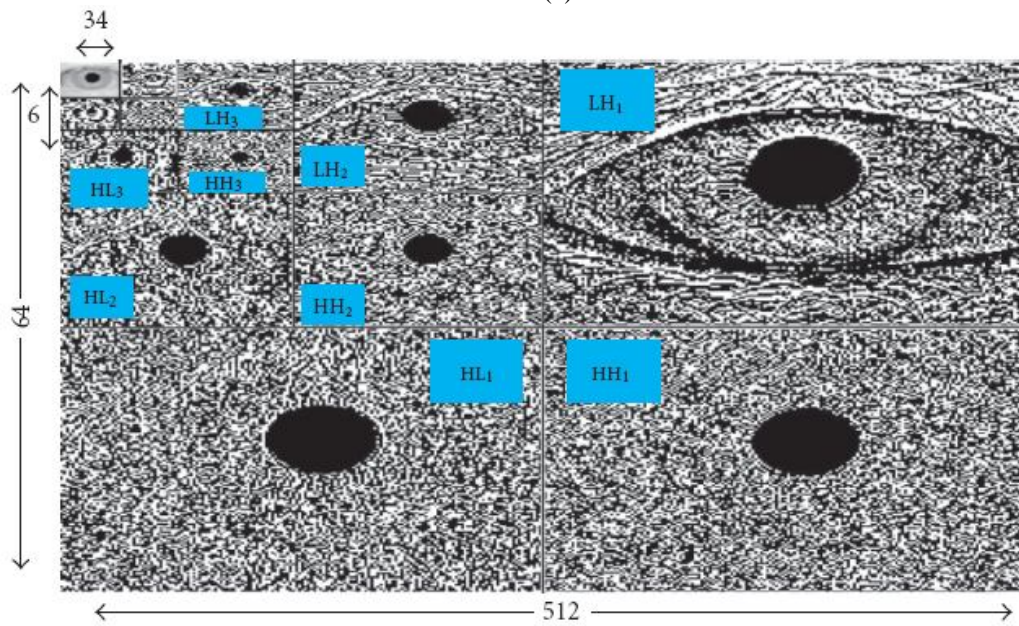
Reference:

- [1] Christel-loïc TISSE¹, Lionel MARTIN¹, Lionel TORRES², Michel ROBERT "Person identification technique using human iris recognition".
- [2] D.E.Benn, M.S.Nixon and J.N.Carter, "Robust eye Extraction using H.T. " AVBPA'99
- [3] D.Gabor, "theory of communication", *journal IEEE*.
- [4] A.Poursaberi and B.N.Araabi, "iris recognition for partially occluded images: methodology and sensitivity Analysis" Hindawi Publishing Corporation EURASIP Journal on Advances in Signal Processing Volume 2007, Article ID 36751, 12 pages...
- [5] [http:// www.neurotechnology.com](http://www.neurotechnology.com)
- [6] John Daugman. "Recognizing persons by their iris patterns "Cambridge University, Cambridge, UK.
- [7] L.M. Waghmare, S. P. Narote, A.S. Narote, "Biometric Personal Identification Using IRIS", Proceedings of International Conference on Systemics, Cybernetics and Informatics, ICSCI - 2006, Pentagon Research Centre – Hyderabad, pp. 679-682, Jan 2006
- [8] E. Wolff, Anatomy of the eye and orbit, H. K. Lewis, London, UK, 7th edition, 1976.

- [9] John Daugman, "How Iris works" *IEEE Transaction on circuit and systems for Video Technology*, VOL.14, No.1, January 2004.
- [10] D.Zang Automated biometrics technologies and systems Klumer Academics, Boston, Mass, USA, 2000.
- [11] A.S. Narote, S.P. narote, M.B. Kokare, L.M. Waghmare, "An Iris Recognition Based on Dual Tree Complex Wavelet Transform" published in the proceedings of IEEE International Conference TENCON 2007, Taiwan during Oct. 30-Nov. 02, 2007.
- [12] R.W.Conners and C.A.Harlow, "A theoretical comparison of texture algorithms", *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol2,no. 3,pp.204-222, 1980.



(a)



(b)

Fig.5 : (a) wavelet decomposition steps diagram and (b) 4-level decomposition of typical image with db2 wavelet

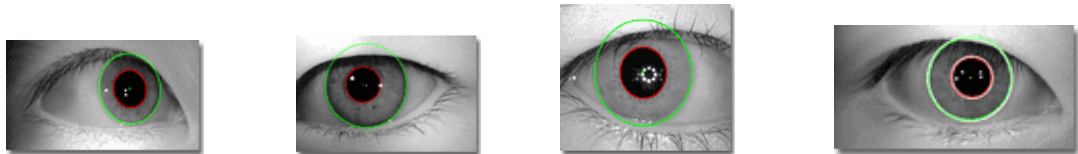


Fig.8

Fig.9

Fig.10

Fig.11

Mn DOPED SnO₂ Semiconducting Magnetic Thin Films Prepared by Spray Pyrolysis Method

K.Vadivel, V.Arivazhagan, S.Rajesh

Abstract -- Semiconducting magnetic thin films of SnO₂ doped with Mn was prepared by spray pyrolysis method. The polycrystalline nature of the films with tetragonal structure was observed from X-ray Diffractometer. The calculated crystalline size was 16-22 nm and the lattice constant is $a=4.73\text{\AA}$ and $c=3.17\text{\AA}$. The compositional studies give the weight percentage of the used materials. The absorption edge starts with 294 nm and rise in transmittance spectra shows the nanocrystalline effect of as deposited films. The calculated band gap from the absorption coefficient is 3.25 eV which greater than the bulk band gap of Tin oxide. The electrical properties of the prepared films also reported in this paper.

Index Terms -- Mn doped SnO₂, Spray Pyrolysis, XRD, UV, Electrical study.

1. INTRODUCTION

THE study of SnO₂ transparent conducting oxide thin films are of great interest due to its unique attractive properties like high optical transmittance, uniformity, nontoxicity, good electrical, low resistivity, chemical inertness, stability to heat treatment, mechanical hardness, Piezoelectric behavior and its low cost. SnO₂ thin films have vast applications as window layers, heat reflectors in solar cells, flat panel display, electro-chromic devices, LEDS, liquid crystal displays, invisible security circuits, various gas sensors etc. Undoped and Cu, Fe and Mn doped SnO₂ thin films have been prepared by vapor deposition technique and reported that SnO₂ belongs to n-type semiconductor with a direct optical band gap of about 4.08 eV [6]. To improve the quality of the films as well as the physical and chemical properties, the addition of some metal ions as

impurities is expected to play an important role in changing the charge carriers concentration of the metal oxide matrix, catalytic activity, the surface potential, the phase composition, the size of crystallites, and so on [8- 10]. It is expected that various concentration of Mn in SnO₂ may affect the structural, optical and magnetic properties of the films. From band gap engineering point of view, suitable band gap is essential for the fabrication of optical devices. So far our knowledge is concerned there are very few reports available on the deposition of Mn doped SnO₂ thin films by spray pyrolysis method. In considering the importance of these materials in the field of magnetic materials, we have prepared Mn doped SnO₂ films using a simple and locally fabricated spray pyrolysis system relatively at the temperature of 450°C.

2. EXPERIMENTAL

Mn doped SnO₂ thin films were prepared by spray pyrolysis method. Mn doped SnO₂ thin films were prepared by spray pyrolysis method. The starting materials were SnCl₄.5H₂O for Tin and Mn(CHOO₃)₂.4H₂O for Manganese. The concentration of 0.5m of Stannous chloride and 0.1m of Manganese acetate was taken in two different beakers with double distilled water. Then 98% of Stannous chloride solution and 2% of manganese acetate solution was mixed together and stirred using magnetic stirrer for 4 hours and allowed to aging for ten days. The clear solution of the mixer was taken for film preparation by spray pyrolysis method. The temperature of the substrate in this method

for preparing nanocrystalline films plays an important role. Here the temperature of the substrate kept at 450°C and the solution was sprayed using atmospheric air as carrier gas. Then the film was allowed to natural cool down. The structural studies on as deposited manganese doped tin oxide thin films were analyzed using X-Ray diffractometer (Shimadzu XRD-6000). Using EDAX (JSM 6390) the elemental composition of the films was carried out. The optical and electrical properties of the films done by UV-Vis spectrometer (Jasco-570 UV/VIS/ NIR) and Hall (Ecopia HMS-3000) measurement system.

3. RESULT AND DISCUSSION

3.1. Structural studies on Mn:SnO₂ thin films

The structural studies on as deposited Mn doped Tin oxide were analyzed by X-Ray spectrophotometer and the graph between 2 theta versus diffracted ray intensity is shown in figure 1. The polycrystalline natures of the prepared samples were observed

from large number of diffracted peaks. The tetragonal structure of the sample with the three strong peaks of (1 1 0), (1 0 1) and (2 1 1) correspond with peak position of 2θ=26.3609, 33.6541 and 51.6145 respectively were identified using standard JCPDS files.

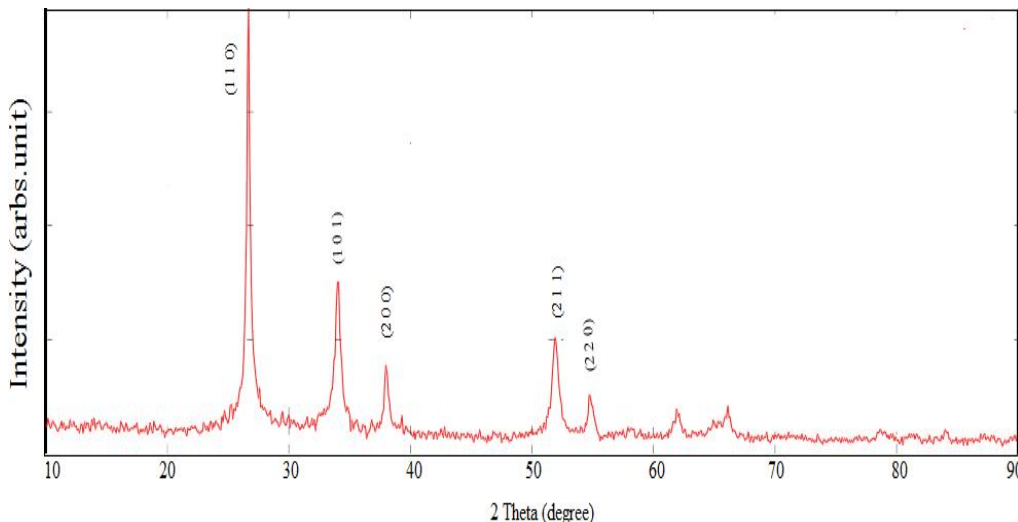


Fig 1. XRD pattern of Mn doped SnO₂ nanocrystalline thin films

Substrate temperature is one of the main parameters, which determine the structural properties of the films. The crystalline size

of as deposited films were calculated using Debye-Scherrer's formula given by,

$$D = \frac{0.9 \lambda}{\beta \cos \theta} \quad \text{----- (1)}$$

Where λ is the wavelength of X-ray used (1.54 Å), β is the full width half maximum (FWHM) of the peak and θ is the glancing

angle. The lattice constant of the spray coated Tin oxide films calculated using the formula

$$\frac{1}{d^2} = \frac{h^2 + k^2}{a^2} + \frac{l^2}{c^2} \quad \text{----- (2)}$$

Where 'd' is the interplanar distance, (h k l) are the Miller indices and 'a' and 'c' are the lattice constant for the Tetragonal structure.

The calculated crystalline size (D) and lattice constant (a and c) of spray coated Mn doped Tin oxide are tabulated in table 1.

Table.1. Structural parameters of Mn doped SnO₂

Substrate temperature	h k l	d (Å)	2θ	FWHM (β)	D (Å)	c (Å)	a (Å)
450° C	1 1 0	3.3420	26.6517	0.3699	220	3.17	4.73
	1 0 1	2.6342	34.0089	0.4690	177		
	2 1 1	1.7608	51.8835	0.5271	167		

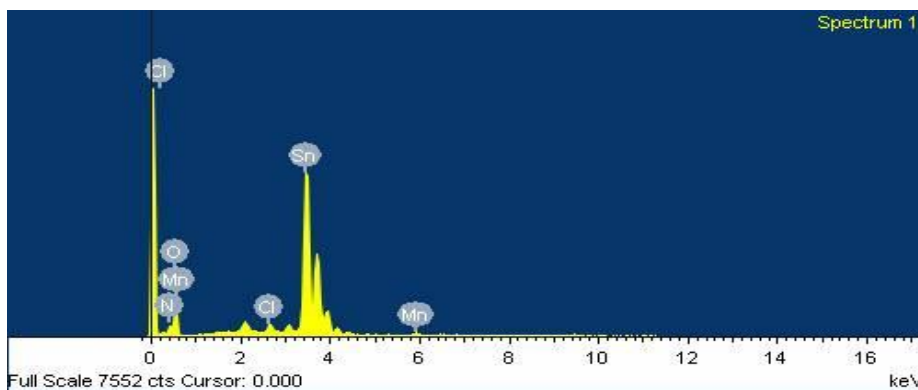


Fig 2. The EDAX spectrum of as deposited Mn doped Tin Oxide thin films

Fig 2 shows, The EDAX spectrum shows the compositional wt % of the used materials. The weight and atomic percentage of Sn was observed as 122.96 % and 23.65 % respectively. The weight and

atomic percentage of doped Mn was observed as 1.09 % and 0.45 % respectively.

3.3. Optical studies on SnO₂ thin film

The optical studies of the Mn doped films were studied by UV –Vis spectrometer in the range of 200-900 nm. The absorption edge starts with 294 nm reveals that the Nanocrystalline effect of

the films. Also the absorption peaks around 400 nm and 550 nm (indicated by arrow) observed in the graph shown in fig 3.

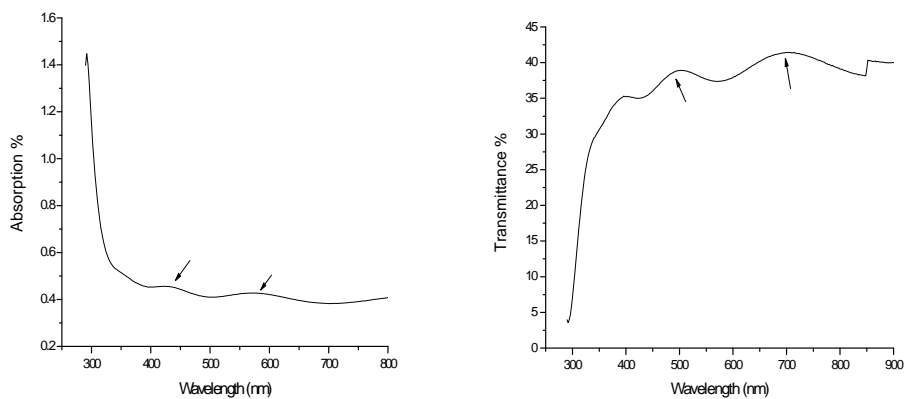


Fig 3. Absorption and transmittance spectra of Mn doped Tin oxide thin films

The light transmitted spectra have the maximum transmittance of 40 percentage is shown in fig 3. The rise in transmittance spectra observed at around 500 nm and 700 nm may be due to

nanocrystalline effect of as deposited films. The optical band gap of the film was calculated using absorption coefficient. The band gap value could be obtained from the optical absorption spectra by using Tauc’s relation [12],

$$\alpha = \frac{A}{h\nu} (h\nu - E_g)^n \text{----- (3)}$$

Where α is the absorption coefficient, $(h\nu)$ is the photon energy and A is a constant. The direct band gap semiconductor can be obtained from the relation,

$$\alpha h\nu = A(h\nu - E_g)^{1/2} \quad (4)$$

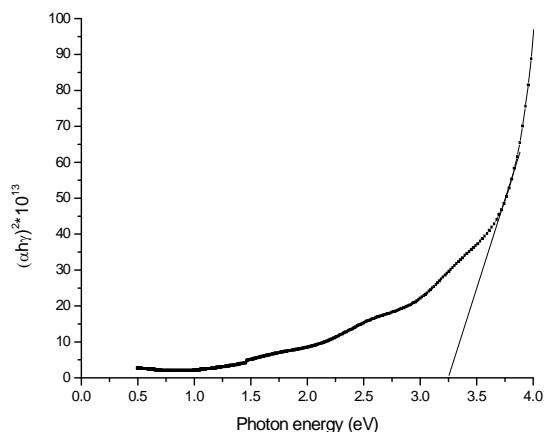


Fig 4. Optical band gap plot between photon energy versus absorption coefficient

Fig.4 shows the variation of $(\alpha h\nu)^2$ versus $(h\nu)$ for the SnO₂ thin film. The straight nature of the films over the wide range of photon energy indicates the direct type of transition. The optical gap has then been determined by extrapolation of the linear region on the energy axis shown in fig.4. The optical band gap of 3.25 eV was

3.3. Electrical properties of Mn doped SnO₂ films

The electrical properties of the prepared films carried out using Hall measurement system at room temperature with the given input voltage of 1 mA. The negative sign of the Hall coefficient value of -3.666×10^{-3} shows the n-type semiconducting nature of the films. The conductivity (σ) and resistivity (ρ) of the film observed

observed in films prepared at 450°C of substrate temperature. It clearly shows the observed value of band gap is greater than the bulk band gap (2.5 eV) of Tin oxide.

as $2.161 \times 10^3 \Omega^{-1} \text{cm}^{-1}$ and $4.628 \times 10^{-4} \Omega \cdot \text{cm}$ respectively. The carrier concentration of the Mn doped SnO₂ have the value of minus $1.703 \times 10^{21} \text{cm}^{-2}$. The mobility of the films were found as $7.922 \text{ cm}^2/\text{V} \cdot \text{sec}$. From these result is observed that the Mn doped SnO₂ films have good electrical properties.

4. CONCLUSION

Manganese doped Tin oxide thin films were prepared by spray pyrolysis method. The X-ray diffractogram shows the polycrystalline nature of as deposited films with tetragonal structure. The crystalline size of the film was calculated using Debye-Scherrer formula is varies from 16-22 nm corresponds to three strong peaks. The calculated lattice constant of the films from interplanar distance and peak plane is $a=4.73\text{\AA}$ and $c=3.17\text{\AA}$.

The optical studies reveals that the presence of nanoparticle on the films. The signature of nanocrystalline effect of as deposited film is absorption edge (294 nm) and the rise in transmittance spectra. The calculated band gap of 3.25 is greater than the bulk band of value of Tin oxide. The n-type semiconducting nature of the films observed from negative sign of the Hall coefficient. The conductivity of $2.161 \times 10^3 \Omega^{-1} \text{cm}^{-1}$ was observed on as deposited films.

REFERNCES

- [1]. Arivazhagan.V , Rajesh.S, Journal of Ovonic research, Vol.6, No.5 ,221-226 ,(2010)
- [2] J. B. Yoo, A. L. Fahrenbruch, R. H. Bube, J Appl Phys. **68**, 4694 (1990).
- [3] R. S. Rusu, G. I. Russia, J. Optoelectron. Adv. Mater **7**(2), 823 (2005).
- [4] M. Penza, S. Cozzi, M. A. Tagliente, A. Quirini, Thin Solid Films, **71**, 349 (1999).

- [5] S. Ishibashi, Y. Higuchi, K. Nakamura, J. Vac. Sci. Technol., **A8**, 1403 (1998).
- [6] J. Joseph, V, K. E. Abraham, Chinese Journal of Physics, **45**, No.1, 84 (2007).
- [7] E. Elangovan, K. Ramamurthi, Cryst. Res. Technol., **38**(9), 779 (2003).
- [8]. Datazoglou O. *Thin Solid Films*, Vol.**302**, 204-213,(1997)
- [9]. Fantini M. and Torriani I. *Thin Solid Films*, Vol.**138**, 255-265 ,(1986).
- [10]. Garcia F.J., Muci J. and Tomar M.S. *Thin Solid Films*, Vol.**97**, 47-51,(1982)
- [11]. Z. C. Jin, J. Hamberg, C. G. Granqvist, J Appl Phys.**64**, 5117 (1988).
- [12]. Advani G.N et al, *Thin Solid Films*, 361 367,(1974)
- [13]. Badawy W.A et al *Electrochem. Soc.*, Vol.**137**, 1592-1595,(1990)
- [14]. Bruneaux J et al,*Thin Solid Films*, Vol.**197**, 129-142,(1991)
- [15]. Chitra Agashe et al. *J. Appl. Phys.* Vol.**70**, 7382-7386,(1991)
- [16]. Chitra Agashe et al, *Solar Energy Mat.*,Vol.**17** ,99-117,(1988)
- [17]. Datazoglou O. *Thin Solid Films*, Vol.**302**, 204-213,(1997)
- [18]. Fantini M. and Torriani I. *Thin Solid Films*, Vol.**138**, 255-265 ,(1986).
- [19]. Garcia F.J., Muci J. and Tomar M.S. *Thin Solid Films*, Vol.**97**, 47-51,(1982)
- [20]. Ghoshtagore R.N. *J. Electrochem. Soc.*, Vol.**125**, 110-17,(1978)
- [21]. Segal and Woodhead *J L Proc.Br.Ceram.Soc.*38, **245**, 1986

*Corresponding Author

K.Vadivel*, V.Arivazhagan, S.Rajesh- Research Department of Physics, Karunya University, Coimbatore, Tamilnadu, India-641 114. *Email: vadivelphyphd08@gmail.com

A Novel comprehensive method for real time Video Motion Detection Surveillance

Sumita Mishra, Prabhat Mishra, Naresh K Chaudhary, Pallavi Asthana

Abstract— This article describes a comprehensive system for surveillance and monitoring applications. The development of an efficient real time video motion detection system is motivated by their potential for deployment in the areas where security is the main concern. The paper presents a platform for real time video motion detection and subsequent generation of an alarm condition as set by the parameters of the control system. The prototype consists of a mobile platform mounted with RF camera which provides continuous feedback of the environment. The received visual information is then analyzed by user for appropriate control action, thus enabling the user to operate the system from a remote location. The system is also equipped with the ability to process the image of an object and generate control signals which are automatically transmitted to the mobile platform to track the object.

Index Terms— Graphic User Interface, object tracking, Monitoring, Spying, Surveillance, video motion detection

1 INTRODUCTION

Video Motion Detection Security Systems (VMDss) have been available for many years. Motion detection is a feature that allows the camera to detect any movement in front of it and transmit the image of the detected motion to the user. VMDss are based on the ability to respond to the temporal and/or spatial variations in contrast caused by movement in a video image. Several techniques for motion detection have been proposed, among them the three widely used approaches are background subtraction optical flow and temporal differencing. Background subtraction is the most commonly used approach in present systems. The principle of this method is to use a model of the background and compare the current image with a reference. In this way the foreground objects present in the scene are detected. Optical flow is an approximation of the local image motion and specifies how much each image pixel moves between adjacent images. It can achieve success of motion detection in the presence of camera motion or background changing. According to the smoothness constraint, the corresponding points in the two successive frames should not move more than a few pixels. For an uncertain environment, this means that the camera motion or background changing should be relatively small. Temporal differencing based on frame difference, attempts to detect moving regions by

making use of the difference of consecutive frames (two or three) in a video sequence.

This method is highly adaptive to dynamic environments hence it is suitable for present application with certain modification. Presently advanced surveillance systems are available in the market at a very high cost. This paper aims at the low cost efficient security system having user friendly functional features which can also be controlled from a remote location. In addition the system can also be used to track the object of a predefined color rendering it useful for spying purposes.

2 HARDWARE SETUP

The proposed system comprises of two sections. The transmitter section consists of a computer, RS232 Interface, microcontroller, RF Transmitter, RF video receiver. The Receiver section consists of a Mobile Platform, RF receiver, microcontroller, RF camera, motor driver, IR LEDs. The computer at the transmitter section which receives the visual information from camera mounted on mobile platform works as control centre. Another function of control centre is to act as the web server that enables access to system from a remote location by using internet. The control centre is also responsible for transmitting the necessary control signal to the mobile platform.

3 MODES OF OPERATION

The system can operate in four independent modes.

3.1 PC Controlled Mode

Sumita Mishra is currently pursuing doctoral degree in Electronics at DRML Avadh University, India and working as a lecturer in electronics and communication engineering department at Amity University, India

E mail: mishra.sumita@gmail.com

Prabhat Mishra is currently pursuing masters degree program in electronics and communication engineering in Amity University, India

In this mode the mobile platform is directly controlled

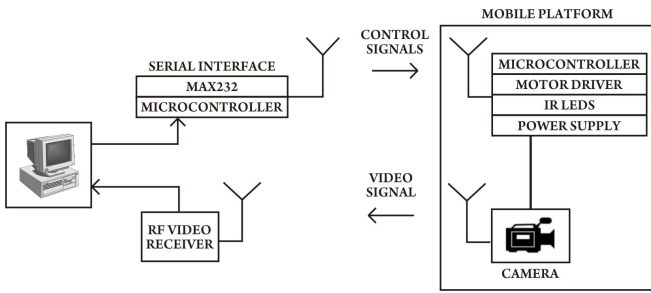


Fig 1. Setup for PC Controlled Mode

by control centre using a visual GUI program developed using Microsoft Visual Studio 6.0(Visual Basic programming language). The user can control the mobile platform after analyzing the received video.

3.2 Internet Controlled Mode

This mode is an extension to the PC Controlled mode where client-server architecture is incorporated. This mode enables an authorized client computer to control the mobile platform from a remote location via internet. Client logs onto the control centre which provides all control tools for maneuvering the mobile platform. Instant images of the environment transmitted from the camera mounted on the mobile

Fig 2. Setup for Internet Controlled mode

platform are used to generate appropriate control signals.

3.3 Tracing Mode

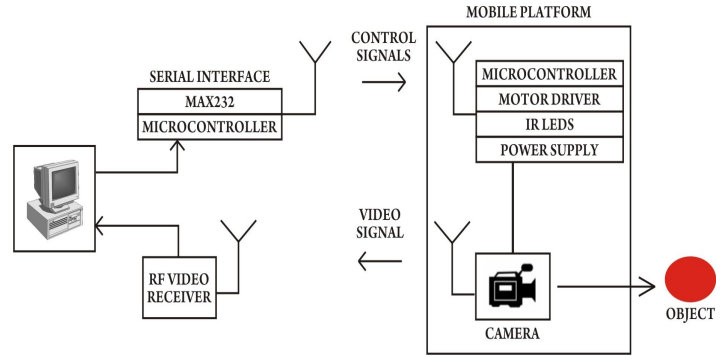
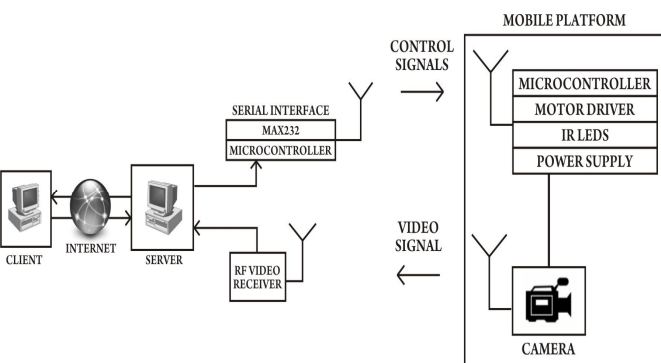


Fig 3. Setup for tracing mode

In this mode the system is made to follow the object whose color information has been stored at the control centre in program developed in MATLAB. Basically the program performs the image processing of the object and generates the control signals in order to make the mobile platform to trace the object.

3.4 Motion Detection Mode

In this mode the platform is made to focus on a particular object whose security is our concern. The mobile platform transmits the visual information of the object to the control centre for analysis. A Program developed using MATLAB at the control centre is then

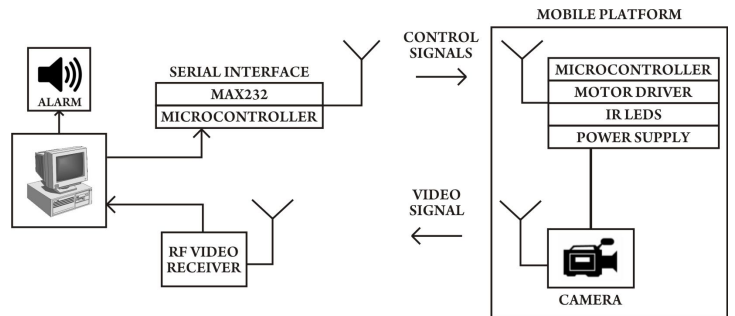


Fig 4. Setup for motion detection mode

used to analyze four consecutive and based on this analysis a security alarm is raised if required.

4. DEVELOPMENT OF CONTROLLING PROGRAMS
4.1 Program for mode 1

This program has been developed in Microsoft Visual Studio 6.0(Visual Basic programming language).It consists of 12 buttons, 2 checkboxes, 1 video box, 1 picture box

box. These 16 buttons are configured as:

- 7 buttons to control the directions of the platform.
- 2 checkboxes for controlling lights and night vision respectively.
- 2 buttons for camera control (start, stop).
- 2 buttons for capturing video.
- 1 button for capturing snapshot

The video box displays the video using VideoCapPro ActiveX Control, received from the camera mounted on the mobile platform and similarly the picture box displays the snapshot taken when the button for capturing the snapshot is depressed. The program transmits the control signals via serial port using MSCOMM (Microsoft Common Control) component.

4.2 Program for mode 2

This program implements client-server architecture in VB using TCP and socket programming.

The client program (fig. 5) has total of 16 buttons,1 video box, 1 picture box and 2 text boxes which are configured as follows:

- 2 buttons for managing the connection between client and server.
- 1 text box to input the host IP address and other for communicating port.
- 7 buttons to control the directions of the mobile platform.
- 2 buttons for controlling lights and night vision respectively.
- 2 buttons for capturing video.
- 1 button for capturing snapshot.
- 2 buttons for camera control (start, stop).

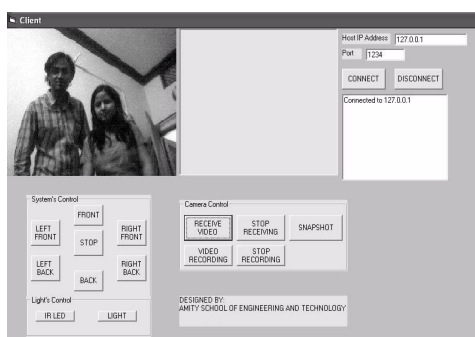


Fig 5. A screen view of client program

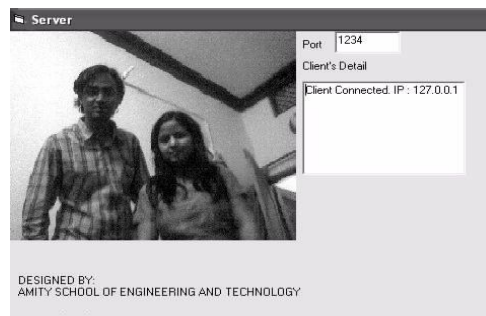


Fig 6 A screen view of server program

The server program (fig. 6) continuously listen to the port and establishes connection if requested. It then transmits the video to the client using VideocapX control and provides access to its serial port.

When connect button is depressed, client program connects to the server whose IP address has been specified in the IP text box at the given port using WINSOCK (Windows Socket) component. Once the connection is established, the server provides video feedback to the client. The client in turn controls the mobile platform via internet.

4.3 Program for mode 3

It has been developed in MATLAB (ver 7.01 from Math Works).The captured image is analyzed pixel by pixel. The screen is divided into four quadrants. Each pixel value is then compared with the stored color value. On comparing these values with the pixels of the captured image (fig. 7) of the object, those pixels are highlighted which matches with the specified color. The highlighted pixels (fig. 8) indicate the direction in which the mobile platform is to be moved. According to the quadrant of highlighted area, control signals are automatically transmitted to the mobile platform to track the object.

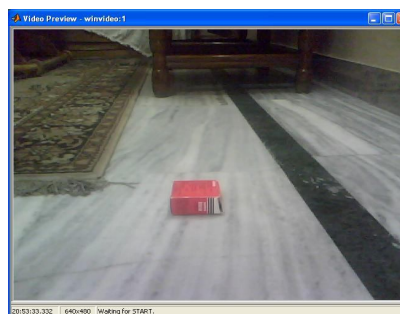


Fig 7 unprocessed image

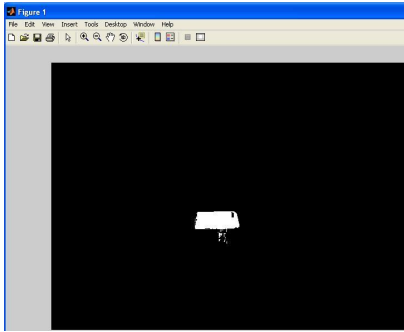


Fig. 8 Processed image showing red color

4.4 Program for mode 4

Again, MATLAB is used for programming. The GUI (fig. 9) has a button and one text box. The button incorporates multiple functions such as start camera, start and stop monitoring process. Monitoring can only be closed if correct password is entered in the text box. On activation the monitoring of scene in front of camera is performed continuously and visual information is sent to the control centre. At the control centre image processing is performed to detect any motion.

4.5 Program for Mobile Platform An AVR ATmega8 microcontroller, L298 motor driver IC is embedded in the mobile platform. The controller is programmed to drive the motor driver IC which in turn drives the motors of the mobile platform. AVR studio ver 4.0 is used as a program editor, WINAVR is used for compiling the code. The program is written into the microcontroller using Ponyprog (ver 1.17) which interfaces the programmer with the parallel port.

5. CONCLUSIONS

The designed advanced real time video motion detection system allows user to maneuver the mobile platform from a remote location. The system also provides feedback to the end user in terms of visuals rendering the system useful for spying purposes. This system integrates the functionality of four modes and selects one of them for unique application. The image processing discussed in mode 3 proves to be useful in industry applications and mode 4 can be useful for monitoring the highly restricted areas. Unlike most previous methods for real time video analysis the suggested approach used in mode 4 utilizes the

difference between four consecutive frames, however, due to the unexpected traffic in the internet, delay in communication between the server and the end user may happen. This causes delays on executing commands and programs and transmitting on-site images.

ACKNOWLEDGMENT

The first author Sumita Mishra is grateful to Prof. B. P. Singh, 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] "An Improved Motion Detection Method for Real-Time Surveillance" Nan Lu, Jihong Wang, Q.H. Wu and Li Yang IAENG International Journal of Computer Science, 35:1, IJCS_35_1_16
- [2] "Robust Salient Motion Detection with Complex Background for Real-time Video Surveillance," Y.L. Tian and A. Hampapur IEEE Computer Society Workshop on Motion and Video Computing, Breckenridge, Colorado, January 5 and 6, 2005.
- [3] "Real-Time Mobile Robot Teleoperation Via Internet Based On Predictive Control", Shihua Wang, Bugong Xu, Yunhui LIU Yeming ZHOU, Front Mech. Engg. China, 3(3):299-306 2008.
- [4] "Internet Control Of a Domestic Robot using a Wireless Lan", Johan Potgeiter, Glen Bright, Olaf Diegel, Sylvester Tiale Mechatronics and Robotic Research Group Institute of Technology and Engineering, Massey University, Auckland
- [5] "Controlling Robot Through Internet Using Java", Mr. Ravindra Thamma, Dr. Luke H. Huang, Dr. Shi-Jer Lou and Dr. C. Ray Diez, Volume 20, Number 3 - June 2004 through August 2004
- [6] <http://gow.epsrc.ac.uk/ViewGrant.aspx?GrantRef=EP/E027253/1>
- [7] <http://www.mathworks.com/matlabcentral/fileexchange/4412>
- [8] http://www.mathworks.com/matlabcentral/newsreader/view_thread/243698

Handoff Analysis for UMTS Environment

Pankaj Rakheja, Dilpreet Kaur, Amanpreet Kaur

Abstract— UMTS is one of the third generation mobile telecommunication technologies. It supports various multimedia applications and services at an enhanced data rate with better security. It also supports mobile users and for that there is a process called handover where new channels are assigned to the user when it moves from a region covered by one node to a region covered by other node. In this paper we are analysing the effect of handover over the performance of the system.

Index Terms— DPCH, Handover, UTRA.

INTRODUCTION

Universal Mobile Telecommunications System [1-2] is a third-generation broadband which supports packet-based transmission of text, digitized voice and video. The multimedia here can reach data rates up to 2 megabits per second (Mbps). It also offers a consistent set of services to mobile computer and phone users, no matter where they are located in the world. It is based on the Global System for Mobile Communications (GSM) standard i.e. it is overlaid on GSM. It is also endorsed by major standard bodies and manufacturers as the planned standard for mobile users around the world. It can ensure a better Grade of Service and Quality of Service on roaming to both mobile and computer users. Users will have access through a combination of terrestrial wireless and satellite transmissions.

Cellular telephone systems used previously [3] were mainly circuit-switched, meaning connections were always dependent on availability of circuits. A packet-switched connection uses the Internet Protocol (IP) [4-5] which uses concept of virtual circuit i.e. a virtual connection is always available to connect an endpoint to the other end point in the network. UMTS has made it possible to provide new services like alternative billing methods or calling plans. For instance, users can now choose to pay-per-bit, pay-per-session, flat rate, or asymmetric bandwidth options. The higher bandwidth of UMTS also enabled other new services like video conferencing. It may allow the Virtual Home Environment to fully develop, where a roaming user can have the same services to either at home, in the office or in the field through a combination of transparent terrestrial and satellite connections.

I. OVERVIEW

The term handover [6] is also known as handoff. Whenever a user terminal moves into area covered by a different RNC while the conversation is still going on, then new channels are allocated to the user terminal which is now under different control node or MSC. This is carried out to ensure continuity of communication and to avoid call dropping. For this to take

place the system needs to identify the user terminal and monitor its signal strength and setting of a threshold value below which a call or application drops and enabling new channel allocation before this level.

There is handoff margin which needs to be optimized for proper synchronization. It is the difference between signal strength at which handover should occur and the minimum required signal strength. If it is too low then there will be insufficient time to complete the process and if it is too large then unnecessary handovers will occur. The most important thing is the handovers are not visible to the users.

Handover types

Handovers can be broadly classified into two types namely: Intracellular and Intercellular handover. In the Intracellular handover, mobile or user terminal moves from one cellular system to another. And in the Intercellular handover, user terminal moves from one cell to the other. This is further classified into soft and hard handover.

Soft handover

Here we follow make before break concept where the user terminal is allocated new channels first and then previous channels are withdrawn. The chances of losing continuity are very less but it needs user terminal or mobile to be capable of toning to two different frequencies. The complexity at user end increases a lot. It is quite reliable technique but here channel capacity reduces.

Hard Handover

Here we follow break before make concept where from the user terminal previously allocated channels are first withdrawn and then new channels are allocated. The chances of call termination are more than in soft handover. At the user terminal complexity is less as it need not be capable of toning to two different frequencies. It provides advantage over soft handover in terms of channel capacity but it is not as reliable as soft handover.

Prioritizing handoffs

Handoff requests are more important than new call requests or application requests as call dropping in between will be more annoying for the user than not being able to make a new call. So a guard channel is especially reserved for the handoffs. We also queue the requests made for proper flow and order control.

The most obvious cause for performing a handover is that due to its movement a user can be served in the another cell more efficiently (like less power emission, less interference etc). It may however also be performed for other reasons which may be system load control.

Classification of cells

Active Set: It is defined as the set of Node-Bs the UE is simultaneously connected to (i.e., the UTRA cells currently assigning a downlink DPCH to the UE constitute the active set).

Monitored Set: It is defined as the set of nodes not in the active set but are included in CELL_INFO_LIST.

Detected Set: It is defined as the set of nodes neither in the active set nor in CELL_INFO_LIST but are detected by UT special considerations in UMTS environment.

In UMTS environment the different types of air interface measurements are:

Intra-frequency measurements: Those measurements which are carried out on downlink physical channels at the same frequency as that of the active set. The measurement object here corresponds to one cell.

Inter-frequency measurements: Those measurements which are carried out on downlink physical channels at frequencies that differ from the frequency of the active set. The measurement object here corresponds to one cell.

Inter-RAT measurements: Those measurements which are carried out on downlink physical channels belonging to another radio access technology than UTRAN, e.g. GSM. The measurement object here corresponds to one cell.

Traffic volume measurements: Those measurements which are carried out on uplink channels to analyse the volume of traffic on them. The measurement object here corresponds to one cell.

Quality measurements: These measurements are carried out on downlink channels to obtain the various quality parameters, e.g. downlink transport block error rate. The measurement

object here corresponds to one transport channel in case of BLER. A measurement object corresponds to one timeslot in case of SIR (TDD only).

UE-internal measurements: Measurements of UE transmission power and UE received signal level.

UE positioning measurements: Measurements of UE position.

The UE supports a number of measurements running in parallel. The UE also supports that each measurement is controlled and reported independently of every other measurement.

II. WORK DONE

We have designed three scenarios [7] where handovers occur when the user terminal moves from the area of one node to the other node. In order to enable communication between source and destination, we made analysis for effect of speed and number of handovers over throughput, average jitter, average end to end delay etc. Here in scenario 1, shown in figure 1, the terrain is 1500 sqms and there are two nodes. The UEs move from the area of one node to that covered by the other node while application is still in active state, so one handover has occurred here; in scenario 2, shown in figure 2, the terrain is 2500 sqms and there are three nodes. The UEs move from the area covered by one node to the second and then to the third one while application is still in active state, so two handovers have occurred here; and in the scenario 3, shown in figure 3, the terrain is 3500 sqms and there are four nodes. The UEs move from the area covered by one node to the second to the third and then to the fourth one while application is still in active state, so three handover have occurred. Here we have taken two users one travelling at 16 m/s and the other at 20 m/s respectively. The later one is called as Fast UT (User Terminal) while the former one is referred to as Slow UT (User Terminal).

The screenshots of three scenarios designed to analyse impact of handover on the overall performance of the system are:

Throughput	Fast UT	Slow UT
One handover	4171	4188
Two handovers	4170	4189
Three handovers	4169	4188

Average Jitter	Fast UT	Slow UT
One handover	0.233	0.213
Two handovers	0.0556	0.009
Three handovers	0.17	0.038

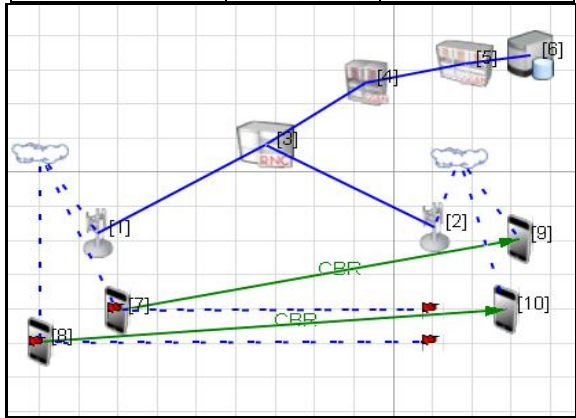


Figure 1: Screenshot of scenario for one handover

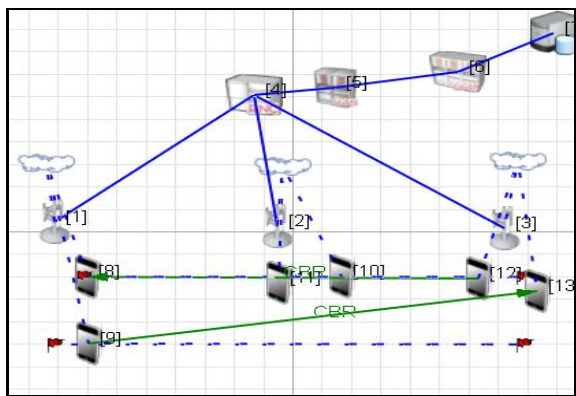


Figure 2: Screenshot of scenario for two handovers

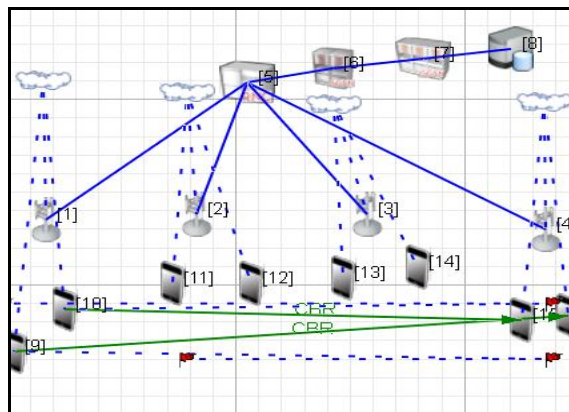


Figure 3: Screenshot of scenario for three handovers

The values obtained by running the simulations for throughput, average jitter and average end to end delay [7] for various scenarios shown in figures (1-3) are:

Table 1: Throughput values for handovers

Table 2: Average jitter values for handovers

Table 3: Average end to end delay values for handovers

Average end to end delay	Fast UT	Slow UT
One handover	0.67517	0.6755
Two handovers	0.45	0.39
Three handovers	0.52	0.44

The plots drawn using these values obtained are:

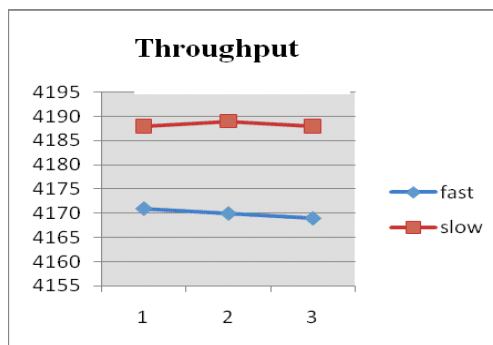


Figure 4: Graph of throughput for handovers

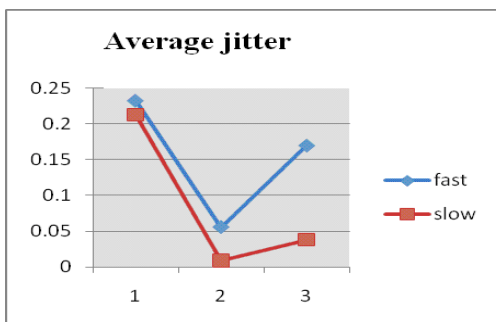


Figure 5: Graph of average jitter for handovers

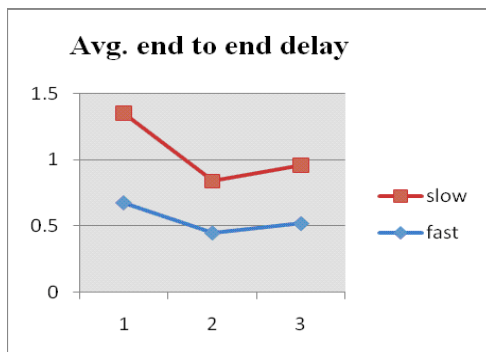


Figure 6: Graph of average end to end delay for handovers

III. CONCLUSION

On analysing the results obtained after running different simulations, we can say that the system has better performance for slow speed users as values of performance determining parameters like throughput, average jitter and average end to end delay are better than that of high speed users. With number of handovers, throughput does not vary much; the average jitter and average end to end delay first fall down and later increases. So, overall performance of system is good in case of handover as average jitter and average end to end delay are not much and throughput is also good as it does not vary much.

IV. REFERENCES

- [1] Falciasacca, G. Frullone, M. Grazioso, P. Riva, G. Serra, A.M., "Performance evaluation of a UMTS in a urban environment", 1991, Sixth International Conference on Mobile Radio and Personal Communications
- [2] van Nielen, M.J.J. ,R. PTT Nederland NV, Leidschendam , "UMTS: a third generation mobile system", Third IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 1992. Proceedings, PIMRC '92

- [3] Oudelaar, j, " Evolution towards UMTS", 5th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 1994
- [4] Fan, L. Sheriff, R.E. Gardiner, J.G," Satellite-UMTS service provision using IP-based technology", 2000 IEEE 51st Vehicular Technology Conference Proceedings.
- [5] Jin Yang Kriaras, "Migration to all-IP based UMTS networks", First International Conference on 3G Mobile Communication Technologies, 2000
- [6] Hyeyeon Kwon Kyung-yul Cheon Aesoon Park," Analysis of WLAN to UMTS Handover", 2007 IEEE 66th Vehicular Technology Conference, 2007
- [7] "Qualnet Programmers Guide", Scalable Network Technologies.

Performance and Emission Characteristics of Stationary CI Engine with Cardnol Bio Fuel Blends

Mallikappa, Rana Pratap Reddy, Ch.S.N.Muthy

Abstract— The compression ignition engine is the most popularly used prime mover. The compression ignition (CI) engine moves a large portion of the world's goods & generates electricity more economically than any other device in their size range [1]. All most all the CI engines use diesel as a fuel, but the diesel is one of the largest contributors to environmental pollution problems. The application of bio diesel as a substitute for conventional petroleum fuel in diesel engine gain ever increasing demand throughout the world wide, because it is produced from renewable resources, bio degradable and potential to exhaust emissions & use of bio diesel in diesel engines generates rural employment opportunities by cultivating such oil producing crops[1-5]. In this research work the detailed investigation on performance and emission characteristics of four stroke single cylinder engine with variable loads were studied, cardnol bio fuel volumetric blends like 0, 10, 15, 20%, and 25% were used. The results indicate that brake power increases (by 76% approximately) as load increases. Brake specific energy conversion decreases (by 30-40 % approximately) with increase in load. Brake thermal efficiency increases with higher loads and emission levels (HC, CO, NOX) were nominal up to 20% blends.

Key words: Compression Ignition, characteristics, cardnol bio fuel, Performance, Emissions

Nomenclature

BSEC	: Brake Specific Energy Consumption
B.T.E	: Brake thermal efficiency
B10	: Blend with 10%bio fuel
CBF	: Cardnol Bio Fuel
CI	: Compression Ignition
CO	: Carbon Monoxide
DR-CNSL	: Double Refined Cashew nut Shell Liquid
EGT	: Exhaust Gas Temperature
HC	: Hydro Carbons
IC	: Internal Combustion
NOx	: Nitrogen oxide
ppm	: Parts per million
Cs	: Centistokes

1 INTRODUCTION

IN today's world the majority of automotive and transportation vehicles are powered by compression ignition engines. The compression ignition engine moves a large portion of the world's goods & generates electricity more economically than any other device in their size range. All most all the CI engines use diesel as a fuel, but the diesel is one of the largest contributors to environmental pollution problems. Bio fuel is an alternative to petroleum based fuel, renewable energy source, bio degradable and non-toxic fuel, being beneficial for reservoirs, lakes, marine life and other environmentally sensitive places such as large cities and mines & use of bio diesel in diesel engines generates rural employment opportunities by cultivating such oil producing crops [1-5].

The issue of energy security led governments and researchers to look for alternate means of renewable and environment-friendly fuels. Bio fuel has been one of the promising, and economically viable alternatives. Fuel and energy crisis and the concern of society for depleting world's non-renewable resources initiate various sectors to look for alternative fuels. One of the most promising fuel alternatives is the vegetable oils and their derivatives.

Plenty of scientific articles and research activities from around the world were printed and recorded. Oils from coconut, soy bean, sunflower, safflower, peanut, linseed and palm were used depending on what country they grow abundantly. It has been reported that in diesel engines; vegetable oils can be used as fuel, straight as well as in blends with the diesel. It is evident that [2] there are various problems associated with vegetable oils being used as fuel in compression ignition engines, mainly caused by their high viscosity. The high viscosity is due to the molecular mass and chemical structure of vegetable oils, which in turn leads the problems in pumping, combustion and atomization in the injector system of diesel engine. Due to the high viscosity, vegetable oils normally introduce the development of gumming, the formation of injector deposits, ring sticking as well as incompatibility with conventional lubricating oils in long-term operations.

India is the largest producer, processor and exporter of Cashews, *Anarcadium Occidentale* Linn, in the world [6]. It was brought to India during the 1400 by Portuguese missionary. Cashew came conquered and took deep root

in the entire coastal region of India. While the tree is native to central and Southern America it is now widely distributed throughout the tropics, particularly in many parts of Africa and Asia. In India Cashew nut cultivation now covers a total area of 0.70 million hectares of land, producing over 0.40 million metric tons of raw Cashew nuts. The Cashew (*Anacardium Occidentale*) is a tree in the flowering plant family Anacardiaceae. The plant is native to northeastern Brazil, where it is called by its Portuguese name Caju (the fruit) or Cajueiro (the tree). It is now widely grown in tropical climates for its cashew "nuts" and cashew apples.

1.1 Specification of Cashew nut shell

The shell is about 0.3 cm thick, having a soft feathery outer skin and a thin hard inner skin. Between these skins is the honeycomb structure containing the phenolic material known as CNSL. Inside the shell is the kernel wrapped in a thin skin known as the testa.

1.2 Composition of cashew nut

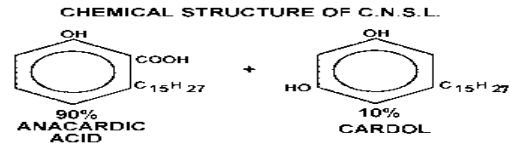
The shell is about 0.3 cm thick, having a soft feathery outer skin and a thin hard inner skin. Between these skins is the honeycomb structure containing the phenolic material known as CNSL. Inside the shell is the kernel wrapped in a thin skin known as the testa. The nut consists of the following kernel 20 to 25%, kernel liquid 20 to 25%, testa 2%, others rest being the shell. The raw material for the manufacture of CNSL is the Cashew.

According to the invention [6] CNSL is subjected to fractional distillation at 200° to 240°C under reduced pressure not exceeding 5mm. mercury in the shortest possible time which gives a distillate containing cardol and the residual tarry matter, for example, in the case of a small quantity of oil, say 200 ml/ the distillation period is about 10 to 15 minutes. A semi-commercial or commercial scale distillation of CNSL may however take longer times. It has been found that there are certain difficulties of operation with regard to single-stage fractional distillation method, i.e. frothing of the oil which renders difficult the fractionation of cardol and also formation of polymerised resin. These difficulties can be overcome in the two-stage distillation, if care is taken not to prolong the heating; this is to avoid the undue formation of polymerised resins and possible destruction partially or completely of the cardol or anacardol. When CNSL is distilled at a reduced pressure of about 2 to 2.5 mm. mercury, the distillate containing anacardol and cardol distills firstly at about 200°C to 240°C. This first distillate is then subjected to a second distillation under the same identical conditions of temperature and pressure when the anacardol distills over at a temperature of 205°C to 210°C and the cardol distills over at a temperature of 230°C to 235°C. In practice it has been found that the preliminary decarboxylation of the oil is essential, since there will be excessive frothing, which renders the distillation procedure unproductive and uneconomical. A specific feature of this invention is that both cardol and anacardol may be obtained by a three-step process. The first step of the process is to get the de-

carboxylated oil by heating the oil to a temperature of 170°C to 175°C under reduced pressure of 30-40 mm. mercury. The next two steps are the same as above for the production of both cardol or cordnol and anacardol.

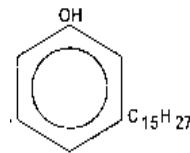
1.2.1 Cardnol

DR-CNSL - Double Refined Cashew nut Shell Liquid. The Cashew Nut Shell Liquid (CNSL) obtained by pyrolysis. It mainly consists two naturally produced phenolic compounds: Anacardic acid 90% Cardol or cardnol 10%.



Cardnol obtained by pyrolysis from dr-csnl oil was utilized for testing purposes. Cardnol is a naturally occurring phenol manufactured from CNSL. It is a monohydroxyl phenol having a long hydrocarbon chain in the Meta position.

$\text{C}_6\text{H}_4(\text{OH})(\text{CH}_2)_7\text{-CH=CH-CH}_2\text{-CH=CH}(\text{CH}_2)_2\text{-CH}_3$



Reason for using the cardnol as alternative fuel: - it is renewable, it is cost effective, easily produced inexpensively in most regions of the world, results in reduced [up to certain extent] emissions compared with petro-diesels, results in no detrimental effects to the engine, non edible and it is extracted from the cashew nut shell not from the seed.

2 EXPERIMENTAL

The main objective was to study the performance and emission characteristics of the CI engine when Cardnol and pure diesel volumetric blends were used and also to investigate which combination of fuel blend is suitable for diesel engine at all load conditions from both performance and emission point of view. Experimentation has been conducted up to cardnol bio fuel volumetric blends like 0, 10, 15, 20%, and 25%, because the viscosities (refer **table 1** for properties of cardnol bio fuel blends) of higher blends are more than the international standard limits [ASTM-allowable limits only up to 4-5 centistokes].

Properties	Diesel	B10	B15	B20	B25	B30
Flash point (C)	50	53	55	56	58	61
Density(Kg/m ³)	817	823	829	836	841	846
Viscosity at 40°C (Centistokes)	2	2.5	3.1	3.5	4.2	5.5
Calorific value (KJ/Kg)	40000	40130	40196	40261	40326	40392

2.2 Properties

Following **table 1** indicates the properties of cardnol bio fuel blends. Lower Calorific value of the diesel has been considered for calculations. **Table 1 Properties**

2.3 Transesterification

Selection of raw materials: cardnol oil sample, anhydrous methyl alcohol 99% grade laboratory reagent type, Sodium Hydroxide was selected as the catalyst.

2.3.1 Procedure

About 4 grams of Naoh (catalyst) is dissolved in 200 ml methanol to prepare alkoxide, which is required to activate the alcohol. Then stirring is done vigorously in a covered container until the alkali is dissolved completely for twenty minutes. Mixture is protected from atmosphere carbon dioxide and moisture as both destroy the catalyst. The alcohol catalyst (Naoh) mixture is then transferred to the reactor containing 700 ml moisture free crude cardnol oil. Stirring of the mixture is continued for 90 minutes at a temperature between 60-65 degrees. The round bottom flask was connected to a reactor condenser and the mixture was heated for approximately three hours.

2.3.2 Inference and observation

The mixture was distilling and condensing within the reactor Condenser, no glycerin, because CNSL is extracted from honeycomb structure (shell) of a cashew nut. The color of cardnol oil slightly changed from dark brown to light brown color and an average of 95% recovery of bio fuel was possible.

2.4 Experimental procedure

The tests were conducted up to 25% blends, because the viscosity of above 25% blends exceeds the international standard limits (i.e. more than 5 Cs). The load test was conducted for different loads i.e. no load, 25%load,50%load, 75%load and full load conditions and for blends such as 0%, 10%, 15%, 20%&25%of Cardnol. The Orotech exhaust gas analyzer used for emission mea-

surements.



Fig 1 Experimental setup

In this investigation the various performance and emission tests were conducted on four strokes single cylinder engine manufactured by M/s Kirloskar (as shown in fig 1) Company limited. The parameter involved in performance analysis has been measured using the engine software supplied by the manufacturer.

2.4.1 Specifications of the engine

Name of the engine: KIRLOSKAR, TV1
General details: 4 stroke, C.I, Vertical
Type of cooling: Water cooled
Number of cylinders: 1
Bore: 87.5 mm
Stroke: 110mm
Rated power: 5.2 B.H.P at 1500 rpm
Dynamometer: Eddy current dynamometer
Compression ratio: 12:1 to 17.5:1

3. RESULTS AND DISCUSSIONS

The experiments were conducted on a direct injection compression ignition engine for various loads with an intention of studying the behavior of the engine in regard to various emissions, and performance characteristics when it was run on different volumetric blends and the results of the performance test and the emission studies conducted on the engine are plotted in the following (Characteristics graphs) figures.

3.1 Brake specific Energy consumption

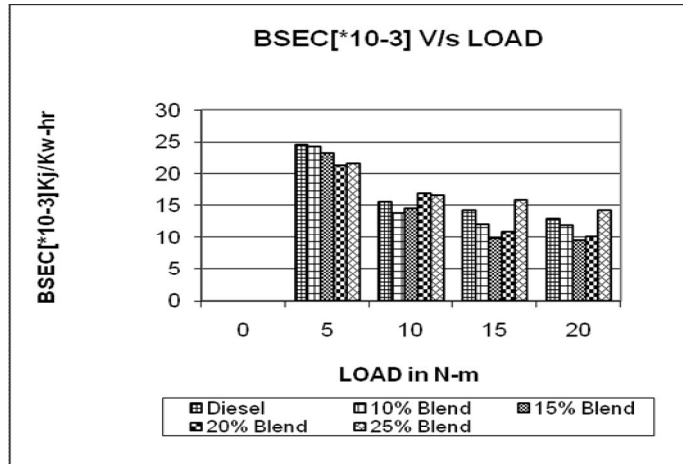


Fig 2. Brake specific Energy consumption (*103Kj/Kw-hr) v/s load

Fig2. Depicts that, the brake specific energy consumption decreases by 30 to 40% approximately with increases in load conditions. This reverse trend was observed due to lower calorific value with increase in bio fuel percentage in the blends.

3.2 Brake thermal efficiency

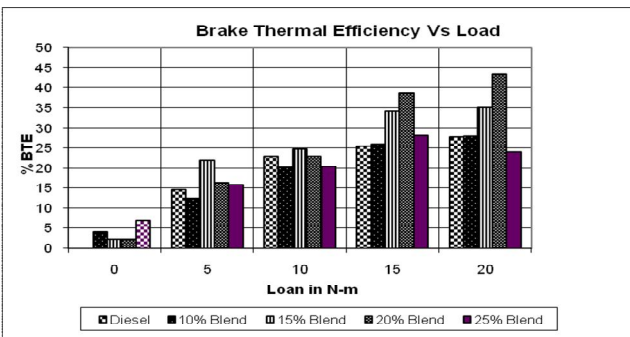


Fig 3. Brake thermal efficiency v/s Load

The variation of brake thermal efficiency with load for different volumetric blends is presented in fig.3. In all cases, it increased with increase in load. This was due to reduction in heat losses and increase in brake power with increase in load. The maximum thermal efficiency for B20 (31%) was higher than that of the diesel. The brake thermal efficiency obtained for B25 was less than that of di-

esel. This lower brake thermal efficiency obtained could be due to lower calorific value and increase in fuel consumption as compared to B20.

3.3 .Exhaust gas temperature & NOx Emissions

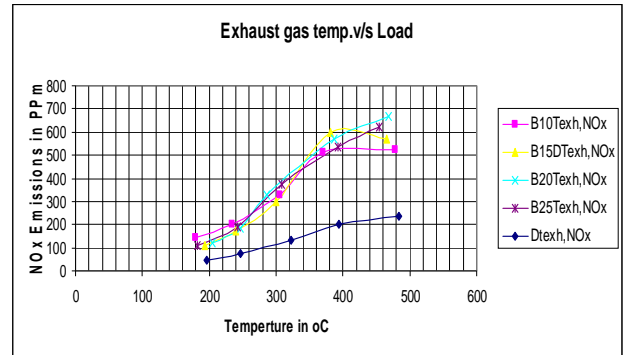


Fig 4. Exhaust gas temp. v/s NOx Emissions

The variations of exhaust gas temperature and Nox emissions with respect to engine loading are presented in the in fig.4. The exhaust gas temperature increases linearly from 180o C at no load to 480 o C at full load conditions. This increasing trend of EGT is mainly because of generating more power and consumptions of more fuel at higher loads.

3.4 HC Emissions

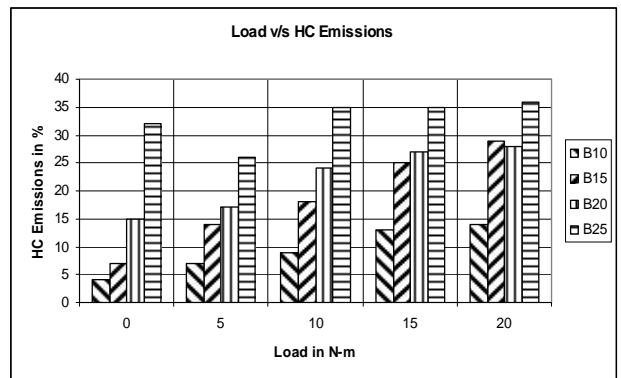


Fig 5. HC Emissions

From the figure 5 it has been observed that HC emissions are nominal up to B20, and more at B25, the reason for this may be incomplete combustion.

3.5 Carbon Monoxide Emissions

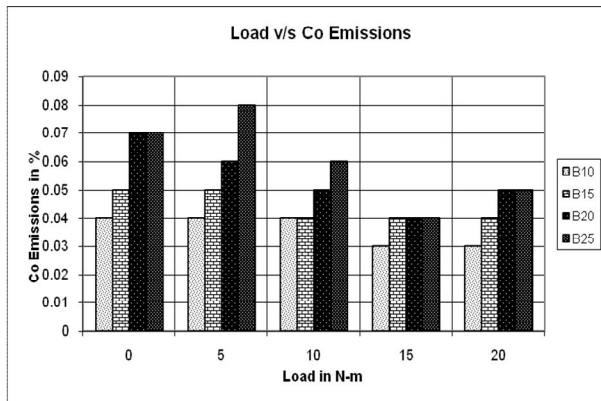


Fig 6. Carbon Monoxide Emissions

Carbon monoxide emissions increases with higher blends, and increases slightly more after 20% blends. The minimum and maximum Co produced was 0.03-0.08%.

At higher loads Co emissions slightly decreased. At elevated temperature, performance of the engine improved with relatively better burning of the fuel resulting in decreased Co.

CONCLUSIONS

The CNSL and its extracts showed promising results in terms of engine performance in par with conventional CI engine fuels. Based on the results of the study the following conclusions were drawn.

The significant factor of cardanol bio fuel is its low cost, its abundance and it is a byproduct of cashew nut industries.

The brake specific energy consumption decreases by 30 to 40% approximately with increases in load conditions. This reverse trend was observed due to lower calorific value with increase in bio fuel percentage in the blends.

The brake thermal efficiency increases with higher loads. In all cases, it increased with increase in load. This was due to reduction in heat losses and increase in brake power with increase in load. The maximum thermal efficiency for B20 (31%) was higher than that of the diesel.

The brake thermal efficiency obtained for B25 was less than that of diesel. This lower brake thermal efficiency obtained could be due to lower calorific value and increase in fuel consumption as compared to B20.

The NOx emissions (ppm) increases with increased proportion of blends and also with higher EGT. This trend mainly because of presence oxygen in bio fuel, this leads to more oxidation at higher temperature and responsible for more NOx emissions.

The HC emissions are nominal up to B20, and more at B25, the reason for this is the incomplete combustion.

The Carbon monoxide emissions increases with higher blends, and increases slightly more after 20% blends. The minimum and maximum Co produced was 0.03-0.08%. At higher loads Co emissions slightly decreased. At elevated temperature, performance of the engine improved with relatively better burning of the fuel resulting in decreased Co.

From this investigation it has been observed that up to 20% blends of cardanol bio fuels may be used in CI engines without any modifications.

ACKNOWLEDGEMENTS

We are very much thankful to M/s Bangalore test house, Bangalore for testing various properties of cardanol oil, & Dr. Sharanappa, Asst. Prof. of Mech. Engg. Dept, and Mr. Viswanath, Instructor R.I.T.M Bangalore, for their kind help during experimentation work. And special thanks Mr. Ganesh Kamath of M/s Sanur Cashew nut Industries, Karkala for supplying CNSL to our experimentation work.

REFERENCES

- [1] Alan C, Lloyd B, Thomas A. Cackette. "Diesel Engines: Environmental Impact and Control" California Air Resources Board, Sacramento, California ISSN1047-3289 J. Air & Waste Manage. Assoc. 51:809-847
- [2] Ayhan Demirbas "Studies on biodiesel from vegetable oils via transesterifications in supercritical Methanol" Energy Conversion and Management 44 (2003) 2093-2109
- [3] Fernando Netoda Silveira*, Antonio Salgado Pratab, Jorge Rocha Teixeiraca "Technical feasibility assessment of oleic sunflower methyl ester utilization in diesel bus engines." Energy conversion and management 44 (2003) 2857-2878
- [4] K. Pramanik "Properties and use of jatropha curcas oil and diesel fuel blends In CI engine," Renewable Energy 28 (2003) 239-248
- [5] N. Stalin and H. J. Prabhu "Performance test of IC engine using Karanja bio diesel blending with diesel" ARPN Journal of engineering and applied science vol.2, no5, October 2007

- [6] Piyali Das, T.Sreelatha, Anurada Ganesh "Bio oil from pyrolysis of cashew nut shell-characterization and related properties "Biomass and bio energy

Morphological Space And Transform Systems

Ramkumar P.B, Pramod K.V

Abstract— Mathematical Morphology in its original form is a set theoretical approach to image analysis. It studies image transformations with a simple geometrical interpretation and their algebraic decomposition and synthesis in terms of elementary set operations. Mathematical Morphology has taken concepts and tools from different branches of Mathematics like algebra (lattice theory), Topology, Discrete geometry, Integral Geometry, Geometrical Probability, Partial Differential Equations etc. In this paper, a generalization of Morphological terms is introduced. In connection with algebraic generalization, Morphological operators can easily be defined by using this structure. This can provide information about operators and other tools within the system

Index Terms—morphological space, transform systems, slope transforms, legendre, kernel.

1 GENERALIZED STRUCTURE FOR MATHEMATICAL MORPHOLOGY

1.1 Definition: Morphogenetic field

Let $X \neq \emptyset$ and $W \subseteq P(X)$ such that i) $\phi, X \in W$,
ii) If $B \in W$ then its complement $\bar{B} \in W$ iii) If $B_i \in W$ is a sequence of signals defined in X , then

$$\bigcup_{i=1}^{\infty} B_i \in W.$$

Let $A = \{ \phi : W \rightarrow U / \phi(\cup A_i) = \vee \phi(A_i) \ \& \ \phi(\cap A_i) = \wedge \phi(A_i) \}$. Then W_u is called Morphogenetic field [14] where the family W_u is the set of all image signals defined on the continuous or discrete images Plane X and taking values in a set U . The pair (W_u, A) is called an operator space where A is the collection of operators defined on X .

1.2 Definition : Morphological Space

The triplet (X, W_u, A) consisting of a set X , a morphogenetic field W_u and an operator A (or collection of operators) defined on X is called a Morphological space. [14]

Example 1. If $X = Z^2$ then it is called Discrete Morphological space

Example 2. Let V be a complete lattice.

If $X = V$ and $A = (V, \vee, \wedge, *, *')$ where $*, *'$ are dilation & erosion then $(V, \vee, \wedge, *, *')$ becomes a commutative complete lattice ordered double monoid or 'Clodum' [11] where $(V, *)$ $(V, *')$ are commutative monoids.

Exmpl 3. If $(V, *)$ $(V, *')$ are groups then $(V, \vee, \wedge, *, *')$ is called a bounded lattice ordered group or blog. [11]

Proposition 1. Clodum is an operator space.

Since Clodum is a particular case as mentioned above, we can consider it as an operator space.

Proposition 2. Blog is an operator space.

Blog is another example for operator space. Similar particular cases exist corresponding to the algebra or geometry under consideration.

1.3 Definition: Dilation

Let (L, \leq) be a complete lattice, with infimum and minimum symbolized by \wedge and \vee respectively. [1] A dilation is any operator $\delta : L \rightarrow L$ that distributes over the supremum and preserves the least element. $\vee_i \delta(X_i) = \delta(\vee_i X_i)$, $\delta(\emptyset) = \emptyset$

1.4 Definition: Erosion

An erosion is any operator $\varepsilon : L \rightarrow L$ that distributes over the infimum [1]. $\wedge_i \varepsilon(X_i) = \varepsilon(\wedge_i X_i)$, $\varepsilon(U) = U$

1.5 Definition: Morphological Adjunctions

Let (X, W_u, A) & (Y, W_u, \bar{A}) be a morphological spaces.

The pair (A, \bar{A}) is called an adjunction iff

$A(X) \leq Y \Leftrightarrow X \leq \bar{A}(Y)$ where \bar{A} is an inverse operator of A .

Proposition 3. Let (X, W_u, δ) & (Y, W_u, ε) be a morphological spaces with operators dilation and erosion on A . Then $\delta(X) \leq Y \Leftrightarrow X \leq \varepsilon(Y)$.

Proposition 4(for lattice). Let (X, W_u, A) & (Y, W_u, \bar{A}) be a morphological spaces. The pair (A, \bar{A}) is called an adjunction iff $\forall u, v \in X, \exists$ an adjunction $(l_{u,v}, m_{v,u})$ on U such

that
$$\bar{A}(x(u)) = \bigvee_{v \in X} m_{v,u}(x(v)) \text{ and}$$

$$A(y(v)) = \bigwedge_{u \in X} l_{u,v}(y(u)), \forall u, v \in X, x, y \in W_U. [1]$$

1.6 Definition: Morphological Kernel

The operator $\phi = \varepsilon \circ \delta$ defines a closure called morphological closure and $\phi^* = \delta \circ \varepsilon$ defines a kernel, called morphological kernel.

2 SLOPE TRANSFORMS - GENERALIZATION

2.1 Introduction

Fourier transforms are most useful linear signal transformations for quantifying the frequency content of signals and for analyzing their processing by linear time – invariant systems. They enable the analysis and design of linear time invariant systems (LTI) in the frequency domain.

Slope transforms are a special type of non linear signal transforms that can quantify the slope content of signals. It provide a transform domain for morphological systems. They are based on eigen functions of morphological systems that are lines parameterized by their slope. Dilation and Erosions are the fundamental operators in Mathematical Morphology. These operators are defined on lattice algebraic structure also. Based on this, Slope transforms are generally divided into three.

They are 1) A single valued slope transform for signals

processed by erosion systems 2) A single valued slope transform for signals processed by dilation systems 3) A multi valued transform that results by replacing the suprema and infima of signals with the signal values at stationary points.

2.2 Special Case – Continuous Time Signals

All the three transforms stated above coincide for continuous-time signals that are convex or concave and have an invertible derivative. This become equal to the Legendre transform (irrespective of the difference due to the boundary conditions).

2.3 Morphological Signal Operators

The morphological signal operators are defined by using Lattice Dilation and Erosion of Signals. The morphological signal operators are parallel or serial inter connections of morphological dilation and erosions, respectively, defined as

$$\begin{aligned} (f + g)(x) &= \bigvee_{y \in R^d} f(x - y) + g(y) \\ \text{and } (f + g)(x) &= \bigwedge_{y \in R^d} f(x + y) - g(y) \end{aligned} \quad (1)$$

Where \bigvee denotes supremum and \bigwedge denotes infimum.

2.4 Legendre Transforms

Let the signal $x(t)$ be concave and assume that there

exist an invertible derivative $x' = \frac{dx}{dt}$. Imagine that the

graph of x , not as a set of points $(t, x(t))$ but as the lower envelope of all its tangent lines. The Legendre transform [12] of x is based on this concept. The tangent at a point $(t, x(t))$ on the graph has slope and intercept equal to $X = x(t) - \alpha(t)$

$\therefore X_L(\alpha) = x[(x')^{-1}(\alpha)] - \alpha(x')^{-1}(\alpha)$ where f^{-1} denotes the inverse.

The function X_L of the tangents intercept versus the slope is the Legendre transform of x [12]

and $x(t) = X_L[(X_L')^{-1}(-t)] + t(X_L')^{-1}(-t)$ If the signal x is convex, then the signal is viewed as the upper envelope of its tangent lines.

2.5 Definition:Upper Slope Transform

For any signal $x: R \rightarrow \bar{R}$ its upper slope transform [12]

is the function $X_\vee: R \rightarrow \bar{R}$ with $X_\vee(\alpha) = \vee_{t \in R} x(t) - \alpha t, \alpha \in R$. The mapping between

the signal and its transform is denoted by $A_\vee: x \rightarrow X_\vee$. If there is one to one correspondence between the signal and its transform, then it is denoted by $x(t) \xrightarrow{A_\vee} X_\vee(\alpha)$.

Proposition6. Upper Slope Transform of a concave signal is equal to its Legendre Transform.

Proof.Let $x(t)$ be a concave signal .Let it has an invertible derivative. For each real α , the intercept of the line passing from the point $(t, x(t))$ in the signals graph with slope α is given by $x(t) - \alpha t$.

For a fixed α , assume that t varies. Let there be a time instant t^* for which the intercept attains its maximum value. The intercept attains its maximum value when the line becomes tangent to the graph. Therefore $x'(t^*) = \alpha$.

Corresponding to the change in α , the tangent also changes, and the maximum intercept becomes a function of the slope α . By its definition, the upper slope transform [12] is equal to this maximum intercept function. Thus, if the signal $x(t)$ is concave and has an invertible derivative, then the upper slope transform is equal to its Legendre transform. Hence the proof.

3 RESULTS BASED ON THE GENERALIZED STRUCTURE

3.1 Definition:Self Conjugate Operator Space

An operator space (W_u, A) is called self conjugate if it has a negation.

Example4. A clodum V has conjugate a^* for every 'a' such that $(avb)^* = a^* \wedge b^*$ and $(a * b)^* = (a^* *' b^*)$ [11]

Example5. If V is a blog [4] then it becomes self conjugate by setting

$$a^* = \begin{cases} a^{-1}, & \text{when } V \text{ inf} < a < V \text{ sup} \\ V \text{ sup}, & \text{when } V \text{ inf} = a \\ V \text{ inf}, & \text{when } V \text{ sup} = a \end{cases} \quad [11]$$

Example6. If X is a concave class then $A^* x(t) = x(-t)$ where $A^* = A \wedge (A \vee)$.

3.2 Definition:Self Conjugate Morphological Space

If the operator space (W_u, A) is self conjugate then the morphological space (X, W_u, A) is called a self conjugate morphological space.

3.3 Definition:Operatable Functions

Let (X, W_u, A) be a morphological space. The collection $K(X, W, A)$ of operatable functions consists of all real valued morphologically operatable functions $x(t)$ defined on X such that $x(t)$ has finite operatability with respect to A . A morphologically operatable function $x \in K$ iff $|x| \in K$.ie. iff $|A(x(\alpha))| \leq A|x(\alpha)|$

3.4 Definition:Morphological Transform Systems

Let (X, W_u, A) be a perfect morphological space and $K = K(X, W_u, A)$ be an operatable space. K is called a morphological transform system if $A[x_T(t)] = X(\alpha) \circ T(\alpha)$

Remark1. Since K is an operatable space,

- 1) $A[x(t) + y(t)] = X(\alpha) + Y(\alpha)$
- 2) $A[x_T(t)] = X(\alpha) \circ T(\alpha)$

3.5 Definition: Morphological Slope Transform System

If $A = A_v$ in the previous definition, then K is called a Morphological slope transform system where A_v is the upper slope transform.

Let (X, W_u, A) be a self conjugate morphological space. If X is a concave class then $A^*(x(t)) = x(-t)$ where $A^* = A \wedge (A \vee)$ and $A \wedge$ is the lower slope transform. Also $A_v(\vee x_c)$

$$= \sum_{\forall c} A_v(x_c)$$

Proposition 7 (Characterization of Slope Transforms).

A Slope transform is an extended real valued function A_v ($A \wedge$) defined on a Morphogenetic field W_u such that

1. $A_v(\phi) = 0$
2. $A_v(x_c) \geq 0 \quad \forall x_c \in W_u$
3. A_v is countably additive in the sense that if (x_c) is any disjoint sequence [or sampling Signal] then $A_v(\vee x_c)$

$$= \sum_{\forall c} A_v(x_c)$$

Remark 2. A_v takes $+\infty$ i.e. $A_v(x_c) = \infty$ if $x(t) = \infty$

$A_v(\alpha) > -\infty, \forall \alpha$ unless $x(t) = -\infty, \forall t$

If $x = \infty$ - then $A_v = -\infty$

Proposition 8. Let K be a morphological transform system.

Let X be a class of concave functions. Let $x(\alpha) \in X$ with each $x(\alpha)$ has an invertible derivative.

Then $A_v(x(\alpha)) = L(x(\alpha))$ where L is the Legendre transform and A_v is the upper slope transform.

4 CONCLUSION

The slope transform has emerged as a transform which has similar properties with respect to morphological signal processing. Fourier transform does this with respect to linear signal processing. Main property of slope transform is that it transforms a supremal convolution (mor-

phological dilation) into an addition. This is similar to the concept in Fourier transform transforms. In Fourier transform a linear convolution changed into a multiplication.

Difference between the Fourier transform and its morphological counterpart, the slope transform is that the Fourier transform is invertible but the slope transform only has an adjoint. In the sense of adjunctions, this means that the 'inverse' of the slope - transformed signal is not the original signal but only an approximation within the sub collection. In this paper we made an attempt for generalizing the algebraic structures related to the theory of Signal processing using Mathematical Morphology. Morphological operators can be redefined by using these structures. We hope that this will be helpful for finding new applications in Mathematical Morphology.

REFERENCES

- [1] John Goustias and Henk J.A.M Heijmans, Mathematical Morphology, I.O.S Press.
- [2] H.J.A.M Heijmans, Morphological Image Operators, Boston, M.A Academic, 1994.
- [3] J. Serra, Image Analysis and Mathematical Morphology, New York Academic, 1982.
- [4] P. Maragos and R.W Schafer, "Morphological system for multi dimensional signal processing" Proc. IEEE, Vol, 78, P.D 690-710, April 1990.
- [5] P. Maragos, A representation theory for morphological image and signal processing. IEEE Transactions on Pattern analysis and machine intelligence 11, (1989), 586-599.
- [6] The Matheron Representation Theorem for Gray Scale Morphological Operators, G. CROMBEZ, Proceedings of the American Mathematical Society Volume 108, Number 3, March 1990 (Proceedings)
- [7] Heijmans, H.J.A.M, and Ronse, C. The algebraic basis of Mathematical Morphology - Part I, Dilations and Erosions, Computer vision, Graphics and Image Processing, 50(1990) 245-295.
- [8] Rein Van Den Boomgaard and Henk Heijmans, Morphological scale space- operators.
- [9] Javier Vidal & Jos'e Crespo, Sets Matching in Binary Images Using Mathematical Morphology, International Conference of the Chilean Computer Science Society.
- [10] Jean Cousty, Laurent Najman and Jean Serra, Some morphological operators in graph spaces, ISSM-2009.
- [11] Petros Maragos, Lattice Image Processing: A Unification of Morphological and Fuzzy Algebraic Systems, Journal of Mathematical Imaging and Vision 22:333-353, 2005.

- [12] Chu-Song Chen, Yi-Ping Hung, Theoretical Aspects of Vertically Invariant Gray-Level Morphological Operators and Their Application on Adaptive Signal and Image Filtering, IEEE transactions on signal processing, vol. 47, no. 4, april 1999 1049.
- [13] Petros Maragos, Slope Transforms: Theory and Application to Nonlinear Signal Processing, IEEE Transactions on Signal Processing, Vol 43, No.4, April 1995.
- [14] K.V Pramod, Ramkumar P.B , Convex Geometry and Mathematical Morphology, International Journal of Computer Applications, Vol:8, Page 40-45.

-
- *Ramkumar P.B is working as Assistant Professor in Mathematics at Rajagiri School of Engineering & Technology, Mahatma Gandhi University, India, PH-04842432058. E-mail: rkpbmaths@yahoo.co.in*
 - *Pramod K.V is working as Professor at Department of Computer Applications, Cochin University, India, PH-01123456789. E-mail :pramodvijayaraghavan@gmail.com*

Bus Proximity Indicator (An Intelligent Bus Stop)

Prof. A.P. Thakare, Mr. Vinod H. Yadav

Abstract— It is always a good idea that a bus commuter waiting at a stop gets to know how far a bus is. If his route of travel happens to be common for more than one bus- route number, it is even better for him to know which is the nearest bus or the earliest arriving bus. This will enable him to opt for the bus or some other mode of commuting. This becomes very useful for the physically challenged commuter, as after knowing in advance the bus arrival s/he will be ready to accommodate in the bus.

A thought of project "Bus Proximity Indicator" is the best solution for the above situation and is best suitable for the B.E.S.T. (The Brihanmumbai Electric Supply & Transport) In this a wireless RF linkage between a certain bus and a bus stop can be used for determination of the bus proximity that helps commuter to know how far his bus is. This project tells him the Bus number, bus name and the approaching time by displaying it on the LCD which is on the bus stop. This project also satisfies the need of automization in bus services.

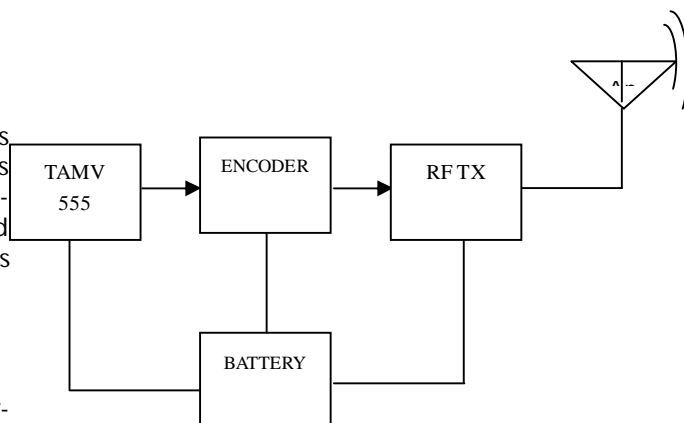
Index Terms— Amplitude Shift Keying, Atmel's AT89C52 Microcontroller, RF encoder/ decoder IC ST12CODEC, C51 Cross Compiler, Radio frequency transmitter, Timer astable multivibrator.

1. INTRODUCTION

THE Bus Proximity Indicator presented in the section uses radio frequency of 433 MHz. The prefixed code of the bus is generated by the encoder ST 12 CODEC. This code is transmitted after Amplitude shift Keying. The receiver positioned at the bus stop detects the radio frequency signal and the bus identification is done by decoder ST 12 CODEC.

2 DESCRIPTION

The block diagram and relevant description of the same is given including of Transmitter and receiving section



2.1 Transmitter Section

The basic block diagram for the Transmitter section is as shown in the block diagram. It consists of the following blocks:

- a) TAMV 555
- b) Encoder
- c) RF Transmitter
- d) Battery

LEGEND:

TAMV – Timer astable multivibrator,
RF TX – Radio frequency transmitter

a) TAMV 555:

The 555 timer IC is used as an astable multivibrator and as an address setter for triggering an IC ST12CODEC which is used as an encoder

(Figure 1: Transmitter Section of Bus Proximity Indicator)

b) RF Encoder:

A logic circuit that produces coded binary outputs from encoded inputs. This uses ST CODEC 12BT for encoding the data. The encoder encodes the data and sends it to RF Transmitter. The IC ST12 CODEC is a single chip telemetry de-

vice, which may be an encoder or a decoder. When combined with a Radio transmitter / receiver it may be used to provide encryption standard for data communication system. The IC ST12CODEC performs all the necessary data manipulation and encryption for an optimum range reliable radio link.

Transmitter and receiver use same IC ST12 CODEC in RF encoder mode for serial communication. This IC is capable of transmitting 12 bits containing 4 bit address bit and 8 bit data. The transmitted information is sent by RF with 434 MHz RF transmitter. ST12 CODEC works on 5v.

RF Transmitter:

RF transmitter's uses **ASK (Amplitude Shift Keying)** for modulating the data send by ST12 CODEC. This modulated information is then transmitted with 433 MHz frequency through RF antenna to receiver. It helps in transmitting data present in encoder via antenna at particular frequency.

c) Battery:

A single 9V battery is used to supply power to the transmitter section.

2.2 Receiver Section

The basic block diagram for the Receiver section is as shown above. It consists of the following blocks,

- a) RF Receiver
- b) RF Decoder
- c) Microcontroller
- d) Power supply
- e) LCD

LEGEND:

RF RX: Radio frequency receiver

LCD: Liquid crystal display

RFDC: RF Decoder

μC: Microcontroller AT 89C51

(Figure 2: Receiver Section of Bus Proximity Indicator)

a) RF Receiver:

It is enhanced single chip IC RWS 434 which receives the 433.92 MHz transmitted signal, transmitted by RF transmitter. It uses ASK (Amplitude Shift Keying) conventional heterodyne receiver IC for remote wireless applications.

b) RF Decoder:

A logic circuit that used to decode coded binary word. This uses IC ST12 CODEC for decoding the data which is transmitted by IC RWS 434. The decoder converts the serial data which has been sent from RF receiver to parallel form and sends it to microcontroller. The coded data decoded by this block is given to LCD.

c) Microcontroller (IC 89C52):

This is the most important block of the entire system. The microcontroller works at crystal frequency of 11.0592 MHz. It receives the parallel data from ST12 CODEC IC and compares it with the program code which already stored in it. This microcontroller has the baud rate 9600 bits/sec.

The 89C52 is a low power, high performance CMOS 8 bit microcomputer with 8k bytes of flash programmable and erasable read only memory (PEROM). The device is manufactured using Atmel's high density nonvolatile memory technology and it is compatible with the industry standard 89C51 and 89C52 instruction set and pin out.

The on chip Flash allows the program memory to be reprogrammed in system or by a conventional nonvolatile memory programmer. By combining a versatile 8-bit CPU with flash on a monolithic chip, the Atmel's AT89C52 is a powerful microcomputer which provides a highly flexible and cost effective solution to many embedded control applications.

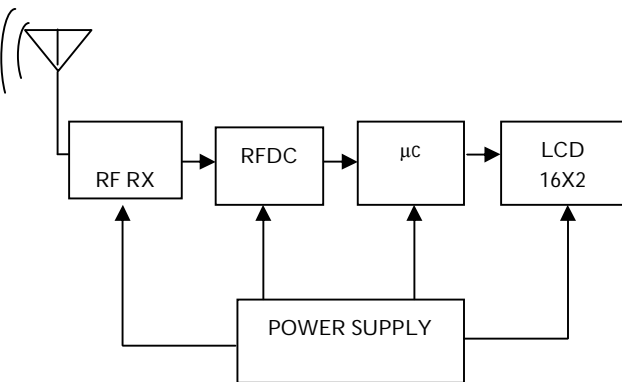
d) Power Supply:

The performance of the master box depends on the proper functioning of the power supply unit. The power supply converts not only A.C into D.C, but also provides output voltage of 5V, 1 amp. The essential components of the power supply are Transformer, four diodes which forms bridge rectifier, capacitor which work as a filter and positive voltage regulator IC 7805. It provides 5v to each block of the transmitter.

e) 16 X 2 LCD:

LCD modules are useful for displaying the information from a system.

These modules are of two types, Text LCD and Graphical LCD. In this project a Text LCD of size (16 x 2) with a two line by sixteen character display is used to display the various sequence of operations during the operation of the project. This is used for visual information purpose. The LCD will display the data coming from normal keyboard or form microcontroller as a visual indication.



Compiler

The Keil C51 Cross Compiler is an ANSI C written specifically to generate fast, compact 8051 microcontroller family. The C51 Compiler generates object code that matches the efficiency and speed of assembly programming.

Software Development Cycle in KEIL

When we use Keil software tools, the development cycle is roughly the same as it is for any software development project.

1. Create a project, select the target chip from the database, and configure the tool settings.
2. Create source file in C or assembly.
3. Build your application with the project manager.
4. Correct errors in the source files.
5. Test the linked application.

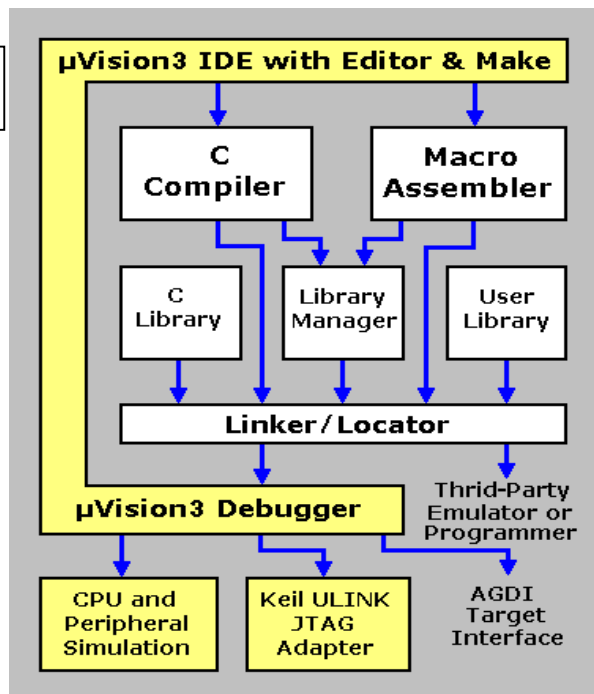
A block diagram of the complete 8051 tool set may best illustrate the development cycle.

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.

4 RESULTS AND CONCLUSION

The paper "**Bus Proximity Indicator**" exhibits the arrival of a particular bus on the display provided at the bus stop. The intention of presenting the paper is to facilitate the commuters waiting at the city bus stops. The CODEC used in the paper generates eight bit coded information allowing the identification up to 256 bus routes and is compatible to the AT89C51. In addition, it works on a power supply ranging from 2 to 5 volts which makes it handy in the mobile

3 SOFTWARE TOOLS



(Figure 3: Software tool of Bus Proximity Indicator) KEIL Introduction to C51 Cross Compiler that is used for the development of any other software from the device

bus. The transmitter and receiver used works on 434MHz at 2 – 12 volt and hence have dual advantage of power saving as well as a range of around 500 feet. The 500 feet (150 Meter) is quite a high range for the detection of the city bus arrival

REFERENCES

- [1] 8051 Microcontroller Architecture Programming and Application by Kenneth J Ayala
- [2] 8051 Microcontroller and Embedded systems by Mazidi and Mazidi
- [3] [Embedded Controller Forth for the 8051 Family](#)
- [4] 8051 Microcontroller, The: Hardware, Software and Interfacing & Applications
- [5] WWW. Google.Com
- [6] [WWW.datasheetcatalog.com](http://www.datasheetcatalog.com)
- [7] digkey.com/1/parts/638617
- [8] <http://www.sunrom.com>
- [9] <http://www.sparkfun.com/products/8946>

-
- Prof. A.P. Thakare Head of Department of Electronics & Telecommunication Sipna's College of Engineering & Technology Amravati 444701 Maharashtra India Email:- asavari_98@yahoo.com
 - Mr. Vinod H. Yadav is currently pursuing masters degree program in Digital Electronics engineering in Sant Gadgebaba Amravati University, Amravati 444701 Maharashtra India, E-mail: anvhie2004@gmail.com, anvhie2004@rediffmail.com, anvhie2004@yahoo.co.in

Designing Aspects of Artificial Neural Network Controller

Navita Sajwan, Kumar Rajesh

Abstract— In this paper important fundamental steps in applying artificial neural network in the design of intelligent control systems is discussed. Architecture including single layered and multi layered of neural networks are examined for controls applications. The importance of different learning algorithms for both linear and nonlinear neural networks is developed. The problem of generalization of the neural networks in control systems together with some possible solutions are also included.

Index Terms— Artificial, neural network, adaline algorithm, levenberg gradient, forward propagation, backward propagation, weight update algorithm.

1 INTRODUCTION

The field of intelligent controls has become important due to the development in computing speed, power and affordability. Neural network based control system design has become an important aspect of intelligent control as it can replace mathematical models. It is a distinctive computational paradigm to learn linear or nonlinear mapping from a priori data and knowledge. The models are developed using computer, the control design produces controllers, that can be implemented online. The paper includes both the nonlinear multi-layer feed-forward architecture and the linear single-layer architecture of artificial neural networks for application in control system design. In the nonlinear multi-layer feed-forward case, the two major problems are the long training process and the poor generalization. To overcome these problems, a number of data analysis strategies before training and several improvement generalization techniques are used.

2 ARCHITECTURE IN NEURAL NETWORKS

Depending upon the nature of the problems, design of neural network architecture is selected. There are many commonly used neural network architectures for control system applications such as Perceptron Network, Adaline network, feed forward neural network.

(a) ADALINE Architecture:

ADALINE (For ADaptive LINear combiner) is a device and a new, powerful learning rule called the widrow-Hoff learning rule this is shown in figure 1. The rule minimized the summed square error during training involving pattern classification. Early applications of ADALINE and its extension to MADALINE (for many ADALINES) include pattern recognition, weather forecasting and adaptive control

ADALINE algorithm:

1. Randomly choose the value of weights in the range -1 to 1.
2. While stopping condition is false, follow steps 3.
3. For each bipolar training pair $S:t$, do step 4-7.
4. Select activations to the input units. $X_0=1$, $x_i=s_i(i=1,2,\dots,n)$.
5. Calculate net input or y .
6. update the bias and weights.
 $W_0=W_0(\text{old})+\alpha(t-y)$
 $W_{\text{new}}=W_i(\text{old})+\alpha(t-y)x_i$
7. If the largest weight change that occurred in step 3 is smaller than a specified value, stop else continue.

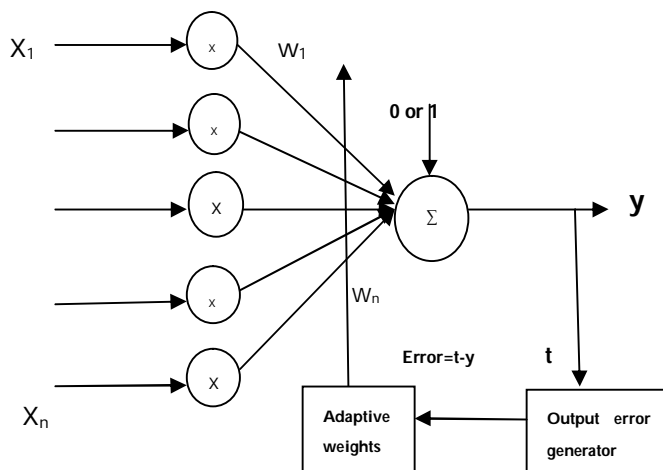


Figure 1: ADALINE Neuron Model

(b) Feed-forward Neural Network Architecture: It is an important architecture due to its non-parametric, non-linear mapping between input and output. Multilayer feed-forward neural networks employing sigmoidal hidden unit activations are known as universal approximators.

These function can approximate unknown function and its derivative. The feed-forward neural networks include one or more layers of hidden units between the input and output layers. The output of each node propagates from the input to the outside side. Nonlinear activation functions in multiple layers of neurons allows neural network to learn nonlinear and linear relationships between input and output vectors. Each input has an appropriate weighting W . The sum of W and the bias B form the input to the transfer function. Any differentiable activation function f may be used to generate the outputs. Most commonly used activation function are purelin $f(x)=x$, log-sigmoid $f(x)=(1+e^{-x})^{-1}$, and tan-sigmoid

$$f(x)=\tan(x/2)=(1-e^{-x}) / (1+e^{-x})$$

Hyperbolic tangent(Tan-sigmoid) and logistic(log-sigmoid) functions approximate the signum and step functions, respectively, and yet provide smooth, nonzero derivatives with respect to the input signals. These two activation function called sigmoid functions because there S-shaped curves exhibit smoothness and asymptotic properties. The activation function f_h of the hidden units have to be differentiable functions. If f_h is linear, one can always collapse the net to a single layer and thus lose the universal approximation/mapping capabilities. Each unit of the output layer is assumed to have the same activation function.

3 BACK PROPAGATION LEARNING

Error correction learning is most commonly used in neural networks. The technique of back propagation, apply error-correction learning to neural network with hidden layers. It also determine the value of the learning rate, η . Values for η is restricted such that $0 < \eta < 1$. Back propagation requires a perception neural network, (no interlayer or recurrent connection). Each layer must feed sequentially into the next layer. In this paper only the three-layer, A, B, and C are investigated. Feeding into layer a is the input vector I . Thus layer a has L nodes, a_i ($i=1$ to L), one node for each input parameter. Layer B, the hidden layer, has m nodes, b_j ($j=1$ to m). $L = m = 3$; in practice $L \neq m$. Each layer may have a different number of nodes. Layer C, the output layer, has n nodes, c_k ($k = 1$ to n), with one node for each output parameter. The interconnecting weight between the i^{th} node of layer A and the j^{th} node of layer B is denoted as v_{ij} , and that between the j^{th} node of layer B and the k^{th} node of layer C is w_{jk} . Each node has an internal threshold value. For layer A, the threshold is T_{Ai} , for layer B, T_{Bi} , and for layer C, T_{ck} . The Back propagation neural network is shown in figure 2..

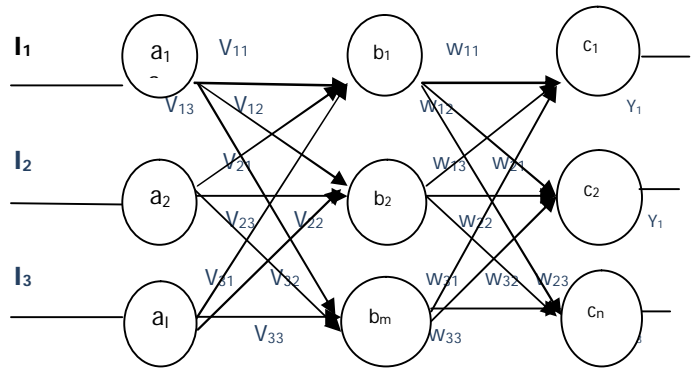


Figure 2: Three-layered artificial neural system

When the network has a group of inputs, the updating of activation values propagates forward from the input neurons, through the hidden layer of neurons to the output neurons that provide the network response. The outputs can be mathematically represented by:

$$Y_p = f\left(\sum_{m=1}^M \left(\sum_{n=1}^N X_n W_{nm}\right) * K_{mp}\right)$$

- Y_p = The pth output of the network
- X_n = The nth input to the network
- W_{nm} = The mth weight factor applied to the nth input to the network
- K_{mp} = The pth weight factor applied to the mth output of the hidden layer
- $F()$ = Transfer function (i.e., sigmoid, etc.)

The ANS becomes a powerful tool that can be used to solve difficult process control applications

Figure 3 depicts the designing procedure of Artificial Neural Network Controller.

4 LEARNING ALGORITHMS

A gradient-based algorithms necessary in the development of learning algorithms are presented in this section. Learning in neural network is known as learning rule, in which weights of the networks are incrementally adjusted so as to improve a predefined performance measure over time. Learning process is an optimization process, it is a search in the multidimensional parameter (weight) space for solution, Which gradually optimizes an objective(cost)function.

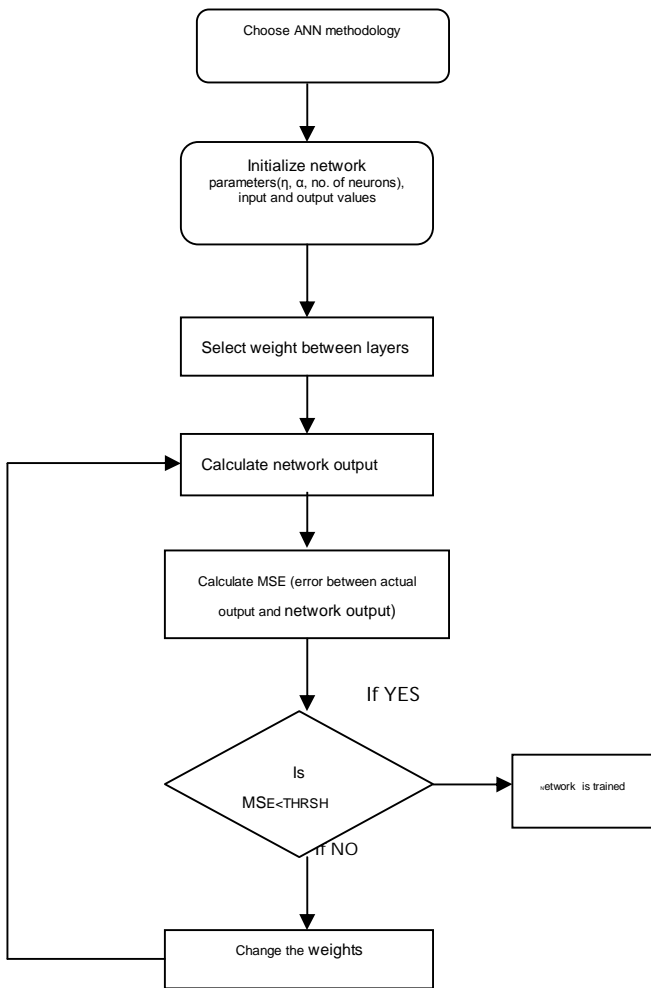


Figure 3: Flow chart for general neural network algorithm

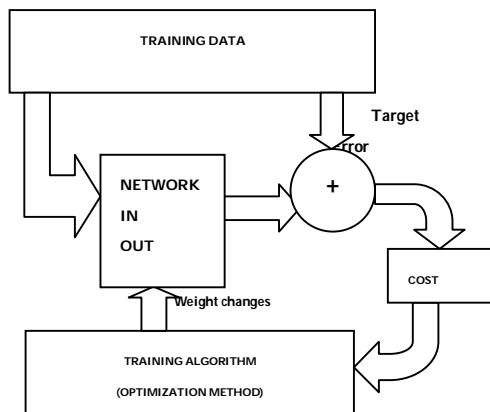


Figure 4: Flow diagram of learning process

5 LEVENBERG GRADIENT BASED METHODS

Gradient based methods searches for minima by comparing values of the objective function $E(\theta)$ at different points. The function is evaluated at points around the current search point and then look for lower values. Objective function $E(\theta)$ is minimized in the adjustable space $\theta = [\theta_1, \theta_2, \dots, \theta_n]$ and to find minimum point $\theta = \theta^*$. A given function E depends on an adjustable parameter θ with a nonlinear objective function. $E(\theta) = f(\theta_1, \theta_2, \dots, \theta_n)$ is so complex that an iterative algorithm is used to search the adjustable parameter space efficiently. The next point θ_{new} is determined by a step down from the current point θ_{now} in the direction vector d given as:

$$\theta_{next} = \theta_{now} + \eta d$$

Where η is some positive step size commonly referred to as the learning rate.

The principal difference between various descent algorithms lie in the first procedure for determining successive directions. After decision is reached, all algorithms call for movement to a minimum point on the line determined by the current point θ_{now} and the direction d . For the second procedure, the optimum step size can be determined by linear minimization as:

$$\eta = \arg(\min(\phi(\eta)))$$

$$\eta > 0$$

where

$$\phi(\eta) = E(\theta_{new} + \eta d).$$

6 LEVENBERG-MARQUARDT METHOD

The Levenberg-Marquardt algorithm can handle ill-conditioned matrices well, like nonquadratic objective functions. Also, if the Hessian matrix is not positive definite, the Newton direction may point towards a local maximum, or a saddle point. The Hessian can be changed by adding a positive definite matrix λI to H in order to make H positive definite.

Thus,

$$\theta_{next} = \theta_{now} - (H + \lambda I)^{-1} g,$$

where I is the identity matrix and H is the Hessian matrix which is given in terms of Jacobian matrix J as $H = J^T J$. Levenberg-marquardt is the modification of the Gauss-Newton algorithm as

$$\theta_{next} = \theta_{now} - (J^T J + \lambda I)^{-1} J^T r.$$

The Levenberg-Marquardt algorithm performs initially small, but robust steps along the steepest descent direction, and switches to more efficient quadratic Gauss-

Newton steps as the minimum is approached. This method combines the speed of Gauss-Newton with the everywhere convergence of gradient descent, and appears to be fastest for training moderate-sized feedforward neural networks.

7 FORWARD-PROPAGATION AND BACK-PROPAGATION

During training, a forward pass takes place. The network computes an output based on its current inputs. Each node i computes a weighted a_i of its inputs and passes this through a nonlinearity to obtain the node output y_i . The error between actual and desired network outputs is given by

$$E = \frac{1}{2} \sum_p \sum_i (d_{pi} - y_{pi})^2$$

where p indexes the pattern in the training set, i indexes the output nodes, and d_{pi} and y_{pi} are, respectively, the desired target and actual network output for the error with respect to the weights is the sum of the individual pattern errors and is given as

$$\frac{dE}{dW_{ij}} = \sum_p \frac{dE_p}{dW_{ij}} = \sum_{p,k} \frac{dE_p}{da_k} \frac{da_k}{dW_{ij}}$$

where the index k represent all outputs nodes. It is convenient to first calculate a value δ_i for each node i as

$$\delta_i = \frac{dE_p}{da_i} = \sum_k \frac{dE_p}{dy_k} \frac{dy_k}{da_i}$$

which measures the contribution of a_i to the error on the current pattern. For simplicity, pattern index p are omitted on y_i , a_i and other variables in the subsequent equations.

For output nodes, dE_p / da_k , is obtained directly as

$$\delta = -(d_{pk} - y_{pk}) f' \quad (\text{for output node}).$$

The first term in this equation is obtained from error equation, and the second term which is

$$\frac{dy_k}{da_k} = f' (a_k) = f'_k$$

is just the slope of the node nonlinearity as its current value. For hidden nodes, δ_i is obtain indirectly as

$$\delta_i = \frac{dE_p}{da_i} = \sum_k \frac{dE_p}{da_k} \frac{da_k}{da_i} = \sum_p \delta_k \frac{da_k}{da_i}$$

where the second factor is obtained by noting that if the node i connects directly to node k then $da_k / da_i = f'_i w_{ki}$, otherwise it is zero. Thus,

$$\delta_i = f'_i \sum_k w_{ki} \delta_k$$

for hidden nodes. δ_i is a weighted sum of the δ_k values of nodes k to which it has connections w_{ki} . The way the nodes are indexed, all delta values can be updated through the nodes in the reverse order. In layered

networks, all delta values are first evaluated at the output nodes based on the current pattern errors, the hidden values is then evaluated based on the output delta values, and so on backwards to the input layer. Having obtained the node deltas, it is an easy step to find the partial derivatives dE_p / dW_{ij} with respect to the weights. The second factor is da_k / dW_{ij} because a_k is a linear sum, this is zero if $k = i$; otherwise

$$\frac{da_i}{dW_{ij}} = X_j$$

The derivative of pattern error E_p with respect to weight w_{ij} is then

$$\frac{dE_p}{dW_{ij}} = \delta_i X_j$$

First the derivative of the network training error with respect to the weights are calculated. Then a training algorithm is performed. This procedure is called back-propagation since the error signals are obtained sequentially from the output layer back to the input layer.

8 WEIGHT UPDATE ALGORITHM

The reason for updating the weights is to decrease the error. The weight update relationship is

$$\Delta w_{ij} = \eta \frac{dE_p}{dW_{ij}} (d_{pi} - y_{pi}) f'_i X_j$$

where the learning rate $\eta > 0$ is a small positive constant. Sometimes η is also called the step size parameter.

The Delta Rule is weight update algorithm in the training of neural networks. The algorithm progresses sequentially layer by layer, updating weights as it goes. The update equation is provided by the gradient descent method as

$$\nabla w_{ij} = w_{ij}(k+1) - w_{ij}(k) = -\eta \frac{dE_p}{dW_{ij}}$$

$$\frac{dE_p}{dW_{ij}} = -(d_{ij} - y_{pi}) \frac{dy_{pi}}{dW_{ij}}$$

for linear output unit, where

$$y_{pi} = \sum_i w_{ij} X_i$$

and

$$\frac{dy_{pi}}{dW_{ij}} = X_i$$

so,

$$\nabla w_{ij} = w_{ij}(k+1) - w_{ij}(k) = \eta (d_{pi} - y_{pi}) X_i$$

The adaptation of those weights which connects the input units and the i th output unit is determined by the corresponding error $e_i = \frac{1}{2} \sum (d_{pi} - y_{pi})^2$.

Training Data Analysis

Two training data analysis methods are:

- (1)normalizing training set and initializing weights
- (2)Principal components analysis(speed up the learning process).

9 CONCLUSION

The fundamentals of neural network based control system design are developed in this paper and are applied to intelligent control of the advanced process. Intelligent control can also be used for fast and complex process control problems.

REFERENCES

- [1] Rajesh Kumar, Application of artificial neural network in paper industry, A Ph.D thesis, I.I.T.Roorkee, 2009.
- [2] S.I.Amari,N.Murata,K.R.Mullar,M.Fincke,H.H.Yang(1997),Asyptotic Statistical Theory of overtraining and cross validation,IEEE Trans.Neural Networks,8(5),985-993.
- [3] C.H.Dagli,M.Akay,O.Ersoy,B.R.Fernandez,A.Smith(1997),Intelligent Engineering Systems Through Artificial Neural Networks,vol.7 of Neural Networks fuzzy logic Data mining evolutionary Programming.
- [4] H.Demuth and M.Beale(1997), Neural Networks toolbox user guide,mathworks.
- [5] L.Fu, Neural Networks in Computer Intelligence(1994).
- [6] M.T.Hang,andM.B.Menhaj(1994),Training feedforward Networks with the Marquardt Algorithm,IEEE Trans. Neural Networks,5(6),989-993.
- [7] Hong HelenaMu,Y.P.Kakad,B.G.Sherlock,Application of artificial Neural Networks in the design of control systems.

Handwritten Character Recognition Using Neural Network

Chirag I Patel, Ripal Patel, Palak Patel

Abstract— Objective is this paper is recognize the characters in a given scanned documents and study the effects of changing the Models of ANN. Today Neural Networks are mostly used for Pattern Recognition task. The paper describes the behaviors of different Models of Neural Network used in OCR. OCR is widespread use of Neural Network. We have considered parameters like number of Hidden Layer, size of Hidden Layer and epochs. We have used Multilayer Feed Forward network with Back propagation. In Preprocessing we have applied some basic algorithms for segmentation of characters, normalizing of characters and De-skewing. We have used different Models of Neural Network and applied the test set on each to find the accuracy of the respective Neural Network.

Index Terms— Optical Character Recognition, Artificial Neural Network, Backpropagation Network, Skew Detection.

1 INTRODUCTION

Such software's are useful when we want to convert our Hard copies into soft copies. Such software's reduces almost 80% of the conversion work while still some verification is always required.

Optical character recognition, usually abbreviated to OCR, involves computer software designed to translate images of typewritten text (usually captured by a scanner) into machine-editable text, or to translate pictures of characters into a standard encoding scheme representing them in (ASCII or Unicode). OCR began as a field of research in artificial intelligence and machine vision. Though academic research in the field continues, the focus on OCR has shifted to implementation of proven techniques [4].

2 ARTIFICIAL NEURAL NETWORK

Pattern recognition is extremely difficult to automate. Animals recognize various objects and make sense out of large amount of visual information, apparently requiring very little effort. Simulating the task performed by animals to recognize to the extent allowed by physical limitations will be enormously profitable for the system. This necessitates study and simulation of Artificial Neural Network. In Neural Network, each node perform some simple computation and each connection conveys a signal from one node to another labeled by a number called the "connection strength" or weight indicating the extent to which signal is amplified or diminished by the connection.

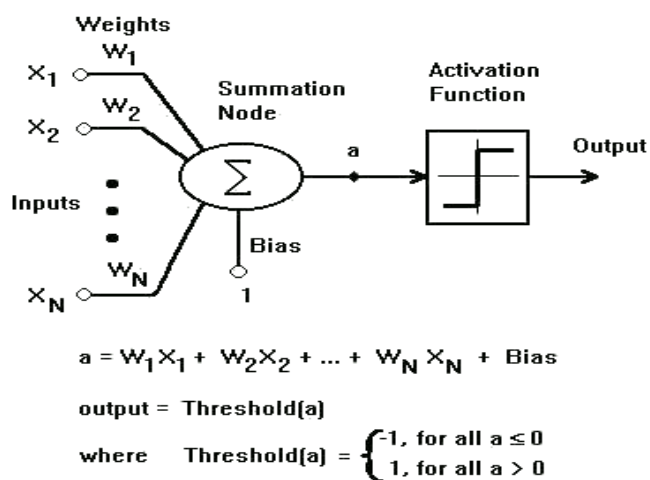


Fig. 1 A simple Neuron

Different choices for weight results in different functions are being evaluated by the network. If in a given network whose weight are initial random and given that we know the task to be accomplished by the network, a learning algorithm must be used to determine the values of the weight that will achieve the desired task. Learning Algorithm qualifies the computing system to be called Artificial Neural Network. The node function was predetermined to apply specific function on inputs imposing a fundamental limitation on the capabilities of the network.

Typical pattern recognition systems are designed using two pass. The first pass is a feature extractor that finds features within the data which are specific to the task being solved (e.g. finding bars of pixels within an image for character recognition). The second pass is the classifier, which is more general purpose and can be trained using a neural network and sample data sets. Clearly, the feature extractor typically requires the most design effort, since it usually must be hand-crafted based on what the applica-

- Chirag I Patel has been completed his M.Tech in Computer Science engineering in Nirma Institute of Technology, Ahmedabad, India, PH-91-9979541227. E-mail: chirag453@gmail.com
- Ripal Patel has been completed her M.E in Electronics & Communication engineering in Dharmsinh Desai Institute of Technology, Nadiad, India, PH-91-9998389428. E-mail: ripalpatel315@gmail.com
- Palak Patel is pursuing her M.E in Electronics & Communication engineering in G.H.Patel College of Engineering & Technology, Vallabh Vidyanagar, India, PH-91-9998389428. E-mail: ipalakec@yahoo.com

tion is trying to achieve.

One of the main contributions of neural networks to pattern recognition has been to provide an alternative to this design: properly designed multi-layer networks can learn complex mappings in high-dimensional spaces without requiring complicated hand-crafted feature extractors. Thus, rather than building complex feature detection algorithms, this paper focuses on implementing a standard backpropagation neural network. It also encapsulates the Preprocessing that is required for effective.

2.1 Backpropagation

Backpropagation was created by generalizing the Widrow-Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer functions. Input vectors and the corresponding target vectors are used to train a network until it can approximate a function, associate input vectors with specific output vectors, or classify input vectors in an appropriate way as defined by you. Networks with biases, a sigmoid layer, and a linear output layer are capable of approximating any function with a finite number of discontinuities.

3 ANALYSIS

By analyzing the OCR we have found some parameter which affects the accuracy of OCR system [1][5]. The parameters listed in these papers are skewing, slanting, thickening, cursive handwriting, joint characters. If all these parameters are taken care in the preprocessing phase then overall accuracy of the Neural Network would increase.

4 DESIGN AND IMPLEMENTATION

Initially we are making the Algorithm of Character Extraction. We are using MATLAB as tool for implementing the algorithm. Then we design neural network, we need to have a Neural Network that would give the optimum results [2]. There is no specific way of finding the correct model of Neural Network. It could only be found by trial and error method. Take different models of Neural Network, train it and note the output accuracy. There are basically two main phases in our Paper: Preprocessing and Character Recognition .

In first phase we have are preprocessing the given scanned document for separating the Characters from it and normalizing each characters. Initially we specify an input image file, which is opened for reading and preprocessing. The image would be in RGB format (usually) so we convert it into binary format. To do this, it converts the input image to grayscale format (if it is not already an intensity image), and then uses threshold to convert this grayscale image to binary i.e all the pixels above certain threshold as 1 and below it as 0.

Firstly we needed a method to extract a given character from the document. For this purpose we modified the graphics 8-way connected algorithm (which we call as EdgeDetection).

5 PREPROCESSING

5.1 Character Extraction Algorithm

1. Create a TraverseList :- List of pixels which have been already traversed. This list is initially empty.
2. Scan row Pixel-by-Pixel.
3. Whenever we get a black pixel check whether the pixel is already in the traverse list, if it is simply ignore and move on else apply Edgedetection Algorithm.
4. Add the List of Pixels returned by Edgedetection Algorithm to TraverseList.
5. Continue the steps 2 - 5 for all rows

5.2 Edge Detection Algorithm

The Edge Detection Algorithm has a list called traverse list. It is the list of pixel already traversed by the algorithm.

EdgeDetection(x,y,TraverseList);

- 1) Add the current pixel to TraverseList. The current position of pixel is (x,y).
- 2) NewTraverseList= TraverseList + current position (x,y).

```
If pixel at (x-1,y-1) then
Check if it is not in TraverseList.
Edgedetection(x-1,y-1,NewTraverseList);
endif
```

```
If pixel at (x-1,y) then
Check if it is not in TraverseList.
Edgedetection(x-1,y,NewTraverseList);
endif
```

```
If pixel at (x-1,y) then
Check if it is not in TraverseList.
Edgedetection(x-1,y+1,NewTraverseList);
endif
```

```
If pixel at (x,y-1) then
Check if it is not in TraverseList.
Edgedetection(x,y-1,NewTraverseList);
Endif
```

```
If pixel at (x,y+1) then
Check if it is not in TraverseList.
Edgedetection(x,y+1,NewTraverseList);
endif
```



```

If pixel at (x+1,y-1) then
Check if it is not in TraverseList.
Edgedetection(x+1,y-1,NewTraverseList);
endif

```

```

If pixel at (x+1,y) then
Check if it is not in TraverseList.
Edgedetection(x+1,y,NewTraverseList);
endif

```

```

If pixel at (x+1,y+1) then
Check if it is not in TraverseList.
Edgedetection(x+1,y+1,NewTraverseList);
endif

```

```

3) return;

```

The EdgeDetection algorithm terminates when it has covered all the pixels of the character as every pixel's position would be in TraverseList so any further call to EdgeDetection is prevented.

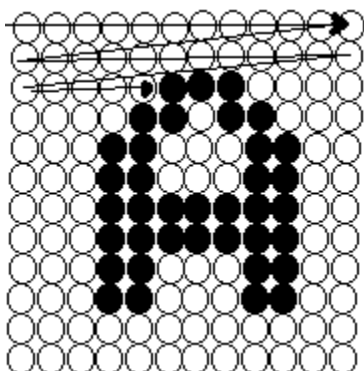
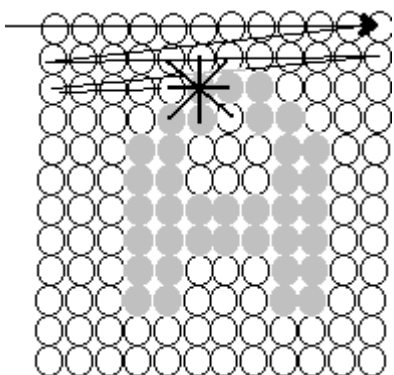


Fig 4(a) shows the traversing each scan lines.



4(b) shows the respective calls made to the all 8-neighbouring pixels.

The Edge detection is called when we hit a pixel (i.e. encounter a pixel with value 1). As per the algorithm the current position is entered in TraverseList and recursive calls are made to all 8 - neighboring pixels. Before the calls are made it is ensured that the corresponding neighboring pixels is having value 1 and is not already encoun-

tered before i.e. it should not be in the TraverseList.

5.3 Normalizing

Now as we have extracted the character we need to normalize the size of the characters. There are large variations in the sizes of each Character hence we need a method to normalize the size.

We have found a simple method to implement the normalizing. To understand this method considers an example that we have extracted a character of size 7 X 8. We want to convert it to size of 10 X 10. So we make a matrix of 70 X 80 by duplicating rows and columns. Now we divide this 70 X 80 into sub Matrix of 7 X 8. We extract each sub matrix and calculate the no. of ones in that sub matrix. If the no. of one's is greater than half the size of sub matrix we assign 1 to corresponding position in normalized matrix. Hence the output would be a 10 X 10 matrix.

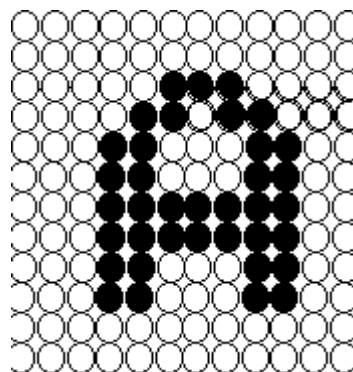


Fig 5(a) shows original representation of the character.

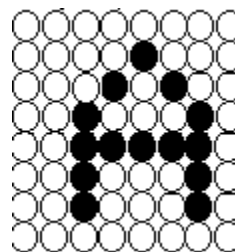


Fig 5(b) shows the Normalized Character representation after Normalizing.

The Fig 5(a) is shows a representation of character of 12 X 12 size. Using the above algorithm it is converted into a character of 8 X 8 as shown in the Fig 5(b).

5.4 Skew Detection

The Characters are often found to be skewed. This would impose problems on the efficient character recognition [3]. So to correct the effect of this skewedness we need counter rotate the image by an angle θ .

We use a very simple but effective technique for Skew Correction. We use "Line Fitting" i.e. Linear Regression to find the angle θ . Consider the Skewed character as a graph i.e. all the pixels that have value 1 are considered to be data points. Then we perform linear regression using the equation $Y = M \cdot X + C$. Using the formulas for regres-

sion we calculate $M = (\sum x_i y_i - \sum x_i \sum y_i) / (\sum x_i^2 - (\sum x_i)^2)$. This angle is equivalent to the skewed angle so by rotating the image by opposite of this angle will remove the skewness. This is a very crude way of removing skewness there are other highly efficient ways of removing skewness. But for Characters that have very low Skew angles this gets the thing done.

The Characters are often found to be skewed. This would impose problems on the efficient character recognition. So to correct the effect of this skewedness we need counter rotate the image by an angle θ .

We use a very simple but effective technique for Skew Correction. We use "Line Fitting" i.e. Linear Regression to find the angle θ . Consider the Skewed character as a graph i.e. all the pixels that have value 1 are considered to be data points. Then we perform linear regression using the equation $Y = M * X + C$. Using the formulas for regression we calculate $M = (\sum x_i y_i - \sum x_i \sum y_i) / (\sum x_i^2 - (\sum x_i)^2)$. This angle is equivalent to the skewed angle so by rotating the image by opposite of this angle will remove the skewness. This is a very crude way of removing skewness there are other highly efficient ways of removing skewness. But for Characters that have very low Skew angles this gets the thing done.

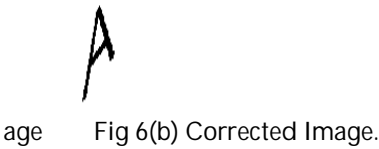


Fig 6(a) Skewed Image

6 NEURAL NETWORK DESIGN

For training and simulating purposes we have scanned certain documents. We have 2 types of documents train documents and test documents. The train documents are the images of the documents which we want to use for training. Similarly test documents are the images of documents which we want to use for test. According to the characters in the documents we train the neural network and apply the test documents.

We have different Models of Neural Network. Hence we record certain parameters like training time, accuracy etc. to find the effectiveness of the Neural Network.

We have selected an image size of 10 X 10 as an input to the Neural Network. Hence we have taken a neural network that has 100 inputs. We are performing the test on only Capital characters so the outputs of the Neural Networks

are 26. The no. of nodes of input layer are 100 and the no. of node of output layer are 26. The no. of hidden layer and the size of hidden layer vary.

Fig 7. The general Model of ANN used.

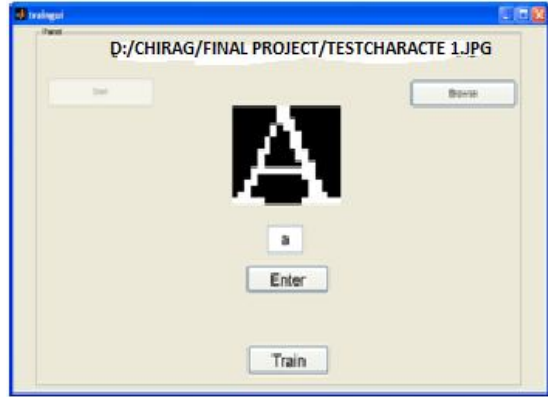


Fig 8. Training automated character extraction

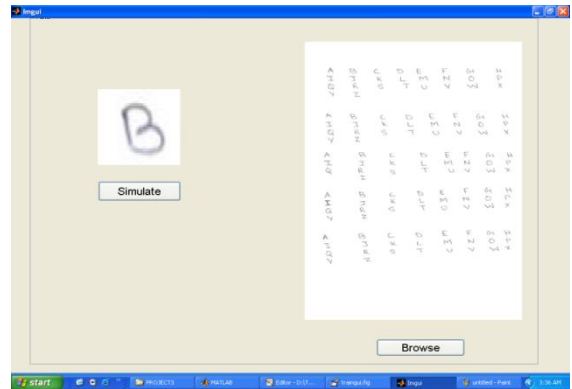


Fig 9 recognition

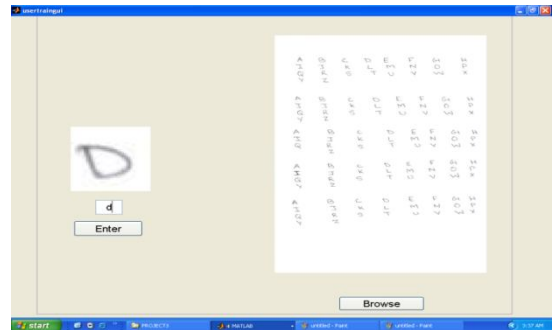
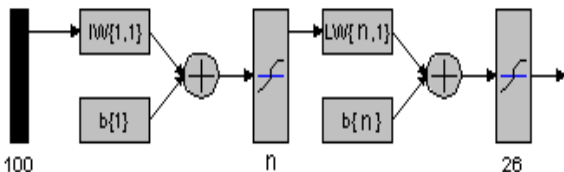


Fig 10 Training – User defined Character Extraction.



7 TEST AND RESULTS ANALYSIS

7.1 Test

This section shows some implementation results. The

training variables involved in the tests were: the number of cycles, the size of the hidden layer, and the number of hidden layer. The dataset consisted of A-Z typed characters of different size and type. Thus the input layer consisted of 100 neurons, and the output layer 26 neurons (one for each character). Ideally, we'd like our training and testing data to consist of thousands of samples, but this not feasible since this data was created from scratch.

Table 1. Model 1

Epochs	Number of Hidden Layer	Configuration (No. of nodes in HL)	(%)
300	1	26	20
600	1	26	65
1000	1	26	82
300	1	52	25
600	1	52	69
1000	1	52	88
300	1	78	27
600	1	78	71
1000	1	78	91

Table 2. Model 2

Epochs	Number of Hidden Layer	Configuration (No. of nodes in HL)	(%)
300	2	26-52	23
600	2	26-52	67
1000	2	26-52	81
300	2	52-78	40
600	2	52-78	78
1000	2	52-78	96
300	2	26-78	27
600	2	26-78	77
1000	2	26-78	89

Table 3. Model 3

Epochs	Number of Hidden Layer	Configuration (No. of nodes in HL)	Accuracy (%)
300	3	26-52-26	31
600	3	26-52-26	65
1000	3	26-52-26	82
300	3	26-52-78	29
600	3	26-52-78	74
1000	3	26-52-78	92
300	3	78-26-78	27
600	3	78-26-78	71
1000	3	78-26-78	91

Table 4. Model 4

Epochs	Number of Hidden Layer	Configur (No. of nodes in HL)	Accuracy (%)
300	3	26-52-26	31
600	3	26-52-26	65
1000	3	26-52-26	82
300	3	26-52-78	29
600	3	26-52-78	74
1000	3	26-52-78	92
300	3	78-26-78	27
600	3	78-26-78	71
1000	3	78-26-78	91

Table 5. Model 5

Epochs	Number of Hidden Layer	Configuration (No. of nodes in HL)	Accuracy (%)
300	4	26-52-78-104	35
600	4	26-52-78-104	79
1000	4	26-52-78-104	96
300	4	26-52-78-26	30
600	4	26-52-78-26	61
1000	4	26-52-78-26	86
300	4	78-26-52-104	43
600	4	78-26-52-104	82
1000	4	78-26-52-104	98
300	4	78-26-78-52	31
600	4	78-26-78-52	88
1000	4	78-26-78-52	94

Table 6. Model 6

Epochs	Number of Hidden Layer	Configuration (No. of nodes in HL)	Accuracy (%)
300	4	26-52-78-104	35
600	4	26-52-78-104	79
1000	4	26-52-78-104	96
300	4	26-52-78-26	30
600	4	26-52-78-26	61
1000	4	26-52-78-26	86
300	4	78-26-52-104	43
600	4	78-26-52-104	82
1000	4	78-26-52-104	98
300	4	78-26-78-52	31
600	4	78-26-78-52	88
1000	4	78-26-78-52	94

We have used sigmoid transfer function in all the layers. We have used same dataset for training all the different Models while testing character set was changed.

7.2 Result Analysis

From the results, the following observations are made:

- A small number of nodes in the hidden layer (eg. 26) lower the accuracy.
- A large number of neurons in the hidden layer help in increasing the accuracy; however there is probably some upper limit to this which is dependent on the data being used. Additionally, high neuron counts in the hidden layers increase training time significantly.
- As number of hidden layer increases the accuracy increases initially and then saturates at certain rate probably due to the data used in training.
- Mostly Accuracy is increased by increasing the number of cycles.
- Accuracy could also be increased by increasing the training set.

7.3 Additional Formatting and Style Resources

Additional information on formatting and style issues can be obtained in the IJSER Style Guide, which is posted online at: <http://www.ijser.org/>. Click on the appropriate topic under the Special Sections link.

8 CONCLUSION

The backpropagation neural network discussed and implemented in this paper can also be used for almost any general image recognition applications such as face detection and fingerprint detection. The implementation of the fully connected backpropagation network gave reasonable results toward recognizing characters.

The most notable is the fact that it cannot handle major variations in translation, rotation, or scale. While a few pre-processing steps can be implemented in order to account for these variances, as we did. In general they are difficult to solve completely.

REFERENCES

- [1] S. Basavaraj Patil, N. V. Subbareddy 'Neural network based system for script identification in Indian documents' in Sadhana Vol. 27, Part 1, February 2002, pp. 83-97.
- [2] T. V. Ashwin, P. S. Sastry 'A font and size-independent OCR system for printed Kannada documents using support vector machines' in Sadhana Vol. 27, Part 1, February 2002, pp. 35-58.
- [3] Kavallieratou, E.; Fakotakis, N.; Kokkinakis, G., 'New algorithms for skewing correction and slant removal on word-level [OCR]' in Proceedings of ICECS '99.
- [4] Simmon Tanner, "Deciding whether Optical Character Recognition is Feasible".
- [5] Matthew Ziegler, "Handwritten Numeral Recognition via Neural Networks with Novel Preprocessing Schemes".

Improved Performance of M-ary PPM in Different Free-Space Optical Channels due to Reed Solomon Code Using APD

Nazmi A. Mohammed, Mohammed R. Abaza and Moustafa H. Aly

Abstract— Atmospheric turbulence induced fading is one of the main impairments affecting the operation of free-space optical (FSO) communication systems. In this paper, the bit error rate (BER) of M-ary pulse position modulation (M-ary PPM) of direct-detection and avalanche photodiode (APD) based is analyzed. Both log-normal and negative exponential fading channels are evaluated. The investigation discusses how the BER performance is affected by the atmospheric conditions and other parameters such as the forward error correction using Reed Solomon (RS) codes and increasing Modulation level. Results strongly indicate that, RS-coded M-ary PPM are well performing for the FSO links as it reduces the average power required per bit to achieve a BER below 10^{-9} in both turbulence channels.

Index Terms— Free Space Optics (FSO), M-ary Pulse Position Modulation (M-ary PPM), Reed Solomon (RS) codes, Log Normal Channel, Negative Exponential Channel, Avalanche Photodiode (APD).

1 INTRODUCTION

Free space optical (FSO) systems have been widely deployed for inter-satellite and deep-space communications. In recent years, however, because of its numerous advantages over radio-frequency (RF) technology such as extremely high bandwidth, license-free and interference immunity, FSO has attracted considerable attention for a variety of applications, e.g. last mile connectivity, optical-fiber backup and enterprise connectivity. In such kind of applications, FSO systems basically utilize atmosphere as transmission medium rather than the free space. So, the performance of FSO link is inherently affected by atmospheric conditions. Among these conditions, atmospheric turbulence has the most significant effect. It causes random fluctuations at the received signal intensity, i.e., channel fading, which leads to an increase in the bit error rate (BER) of the optical link [1].

Current FSO communication systems employ intensity modulation with direct detection (IM/DD) and use light emitting diodes (LED) or laser diodes as transmitters and PIN photodiode or avalanche photodetectors (APD) as receivers. These devices modulate and detect solely the intensity of the carrier and not its phase. Furthermore, biological safety reasons constrain the average radiated optical power, thereby constraining the average signal amplitude. The most reported modulation technique used

for FSO is the on-off keying (OOK) which offers bandwidth efficiency but lacks power efficiency. Binary level signaling though is the simplest and most common modulation scheme for the optical intensity channel and offers low power efficiency and high bandwidth efficiency. Power efficiency as well as the improved system performance can be achieved by adopting pulse position modulation (PPM) schemes. M-ary PPM achieves high power efficiency at the expense of reduced bandwidth efficiency compared with other modulation schemes. The optimal PPM order is high, since a higher order modulation creates the higher peak power needed to overcome the weak average power. M-ary PPM has been previously suggested as a suitable modulation scheme for FSO systems [2]. The IrDA specification for the 4 Mbps short distance wireless infrared links specifies a 4-PPM modulation scheme [3].

Reed Solomon (RS) codes are a class of block codes that operate on symbols rather than bits. So, RS codes can correct both random bit errors and burst symbol errors. Moreover, their hard decoding algorithm can be easily implemented even at a high operation speed. International Telecommunication Union-Telecommunication (ITU-T) standard forward error correction (FEC) scheme based on RS (255,239) codes has been widely used in 10 Gbps practical optical fiber transmission systems [4]. Kiasaleh derived upper bounds on the BER of M-ary PPM over log-normal and exponential distributed channels, when an APD is used [5].

In this paper, Kiasaleh expressions are used to compare the effect of the BER of average photons per PPM bit with the former channels without coding and with RS (255,207) coding. The remainder of the paper is organized as follows. The models of FSO channels are presented in Section 2. Based on the theory presented, a numerical

- Nazmi A. Mohammed is with the Electronics and Communications Engineering Department, Arab Academy for Science, Technology and Maritime Transport, Egypt. naz_azz@yahoo.com.
- Mohammed R. Abaza is with the Electronics and Communications Engineering Department, Arab Academy for Science, Technology and Maritime Transport, Egypt. OSA Student Member. rauf_abaza@hotmail.com.
- Mostafa H. Aly is with the Electronics and Communications Engineering Department, Arab Academy for Science, Technology and Maritime Transport, Egypt. OSA Member. drmosaly@gmail.com.

analysis of the M-ary PPM is carried out in Section 3. This is followed by the main conclusions in Section 4.

2 MODELS OF FSO CHANNELS

2.1 Log-Normal Channel

The log-normal channel is classified as “weak turbulence”, which is characterized by a scintillation index less than 0.75. In general, the scintillation index is a complicated function of the beam parameters, propagation distance, heights of the transmitter and receiver, and the fluctuations in the index of refraction. In fact, the main source of scintillation is due to fluctuations (due to temperature variations) in the index of refraction, which is commonly known as optical turbulence. The log-normal model is also valid for propagation distances less than 100 m [5].

The bit error rate (BER) of an M-ary PPM in log-normal channel is given by [6]

$$P_b^M \leq \frac{M}{2\sqrt{\pi}} \sum_{i=-N, i \neq 0}^N w_i Q \left(\sqrt{\frac{e^{2(\sqrt{2}\sigma_k x_i + m_k)}}{Fe^{\sqrt{2}\sigma_k x_i + m_k} + K_n}} \right) \quad (1)$$

where M is the modulation level, w_i and x_i are the weight factors and the zeros of the Hermite polynomial [7].

The scintillation index (σ_{SI}^2) as a function of the variance of the log-normal channel (σ_k^2) is given by [5]

$$\sigma_{SI}^2 = e^{\sigma_k^2} - 1 \quad (2)$$

The average photons per PPM slot ($E\{K_s\}$) are functions of the mean (m_k) and the variance of the log-normal channel and have the form [5].

$$E\{K_s\} = e^{\left(\frac{\sigma_k^2 + m_k}{2}\right)} \quad (3)$$

The total noise photons per PPM slot, K_n , which results from background noise and thermal noise, is [5]

$$K_n = \frac{2\sigma_n^2}{(E\{g\}q)^2} + 2FK_b \quad (4)$$

where K_b is the average background noise photons per PPM slot, $E\{g\}$ is the average gain of the APD and q is the electron charge.

The noise factor, F , of the APD is defined by [5]

$$F = 2 + \zeta E\{g\} \quad (5)$$

where ζ is the ionization factor.

The variance, σ_n^2 , of the thermal noise in a PPM slot is defined by [5]

$$\sigma_n^2 = \left(\frac{2KT T_{slot}}{R_L} \right) \quad (6)$$

where T is the effective absolute temperature of the receiver, K is Boltzmann constant, R_L is the APD load resistance and T_{slot} is the PPM slot duration which is related to the data rate, R_b , by [5]

$$T_{slot} = \frac{\log_2(M)}{MR_b} \quad (7)$$

In case of coding, R_b must be multiplied by (n/k) , where n is the codeword length and k is the message length.

The symbol error rate (P_{symbol}) can be calculated from bit error rate (P_b) as [8]

$$P_{symbol} = P_b \left(\frac{2(M-1)}{M} \right) \quad (8)$$

The probability of the uncorrectable symbol error (P_{ues}) due to RS codes can be calculated by the formula [9, 10]

$$P_{ues} \leq \frac{1}{n} \sum_{i=t+1}^n i \binom{n}{k} P_q^i (1 - P_q)^{n-i} \quad (9)$$

where $t = ((n-k)/2)$ is the symbol error correcting capability, P_q is the q-bit RS symbol error probability.

The BER after coding (P_{bc}) is given by [9]

$$P_{bc} = P_{ues} \left(\frac{n+1}{2n} \right) \quad (10)$$

2.2 Negative Exponential Channel

The negative exponential channel is classified as “strong turbulence”, which is characterized by a scintillation index greater than 1. The negative exponential model is valid for propagation distances more than 100 m or several kilometers [5, 8].

The BER of the negative exponential channel, P_b^M , is given by [5].

$$P_b^M \leq \frac{M}{2} \sum_{i=-N, i \neq 0}^N w_i |x_i| Q \left(\frac{E\{K_s\} x_i^2}{\sqrt{FE\{K_s\} x_i^2 + K_n}} \right) \quad (11)$$

To get the BER after coding (P_{bc}) due to negative exponential channel using RS (255,207), we apply the same procedure of the log-normal channel coding steps, which are

mentioned in (7), (8), (9), (10) and use this in (11).

3 NUMERICAL RESULTS AND DISCUSSIONS

Based on the described model, the BER of 10^{-9} , which is considered as a practical performance target for FSO link [11], is calculated for log-normal channel and negative exponential channel and the obtained results are displayed in Figs. 1-3. In these figures, the value of scintillation index (σ^2_{si}) is taken 0.3 for weak turbulence and 1 for strong turbulence. The values of other parameters are taken as: BER=2.4 Gbps, $K_b = 10$ photons per PPM slot, $RL=50 \Omega$, $\varsigma=0.028$, $T=300^\circ K$, $E\{g\}=150$, $n=255$, $k=207$, $t=24$ symbols.

Variations of BER with the average number of photons received per PPM bit (logarithmic), which are equal $\log_{10}(E\{K_s\}/\text{number of bits in PPM symbol})$, are shown in the following figures. In discussions, all results of average photons per PPM bit are in numerical values.

In Fig. 1, binary PPM (BPPM) is used to compare between the effect of weak turbulence and strong turbulence without coding and with RS (255,207) coding.

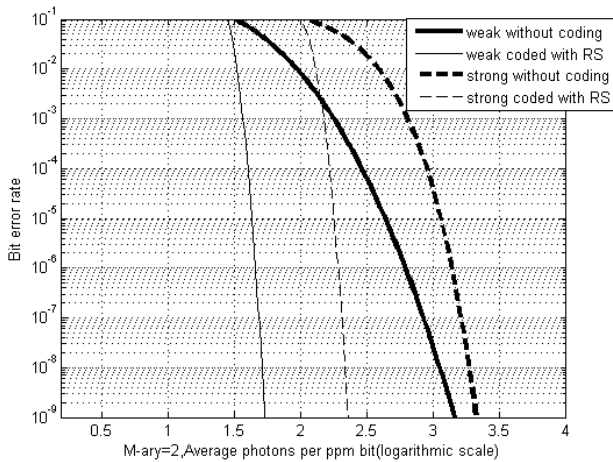


Fig.1. Comparison of non-coded BPPM in weak turbulence of 0.3 scintillation index with Reed Solomon (255,207) and in strong turbulence of 1.0 scintillation index.

At a BER of 10^{-9} , the value of average photons per PPM bit in strong turbulence is found 2046 without coding and 219 with coding which gives an improvement of 9.7 dB. While the value of average photons per PPM bit in weak turbulence is found 1462 without coding and 52 with coding which gives an improvement of 14.49 dB. It also shows that, coded strong turbulence BPPM outperforms weak turbulence BPPM without coding at a BER less than 10^{-3} .

In Fig. 2, 8-PPM is used to show the effect of multilevel modulation and to compare between the effect of weak turbulence and strong turbulence without coding and with RS (255,207) coding.

At a BER of 10^{-9} , the value of average photons per PPM bit in strong turbulence is found 728 without coding and 141 with coding giving an improvement of 11.61 dB.

While the value of average photons per PPM bit in weak turbulence is found 543 without coding and 32 with coding which gives an improvement of 16.6 dB. It is shown that, coded strong turbulence 8-PPM outperforms weak turbulence 8-PPM without coding at BER less than 10^{-4} .

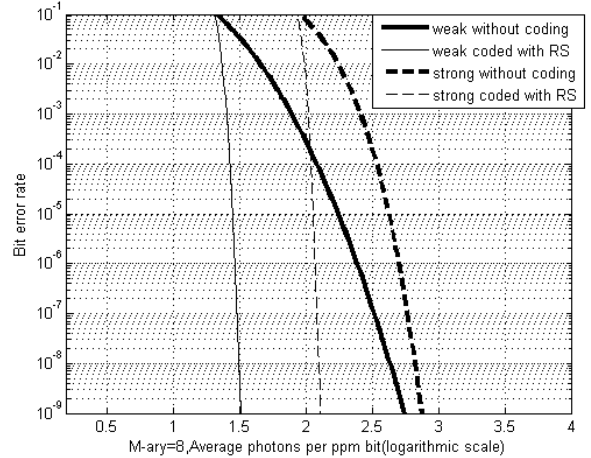


Fig. 2. Comparison of non-coded 8-PPM in weak turbulence of 0.3 scintillation index with Reed Solomon (255,207) and in strong turbulence of 1.0 scintillation index.

In Fig. 3, 256-PPM is used for weak turbulence and strong turbulence without coding and with RS (255,207) coding.

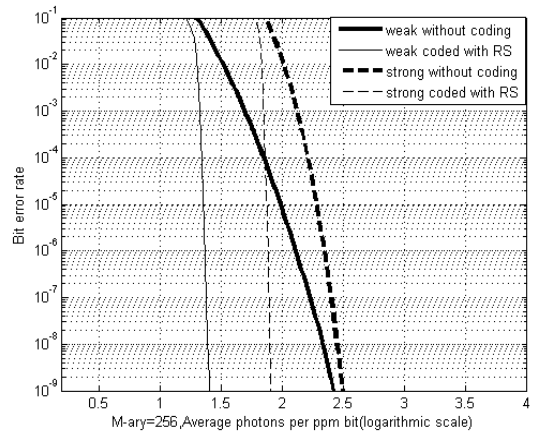


Fig. 3. Comparison of non-coded 256-PPM in weak turbulence of 0.3 scintillation index with Reed Solomon (255,207) and in strong turbulence of 1.0 scintillation index.

At a BER of 10^{-9} , the value of average photons per PPM bit in strong turbulence is found 308 without coding and 80 with coding showing an improvement of 14.08 dB. While, in weak turbulence, the value of average photons per PPM bit is found 258 without coding and 25 with coding giving an improvement of 17.67 dB. It also shows that, coded strong turbulence 256-PPM outperforms weak turbulence 256-PPM coding less at BER less than 10^{-4} .

The obtained results for coded and non-coded M-ary PPM are summarized and compared in Tables 1 and 2 for strong and weak turbulence, respectively.

TABLE 1

Average number of photons per ppm bit required achieving 10^{-9} in strong turbulence.

M-ary PPM	Strong turbulence	Strong turbulence with RS code	Improvement due to RS code + multilevel modulation in Strong turbulence respect to non coded BPPM	
			%	dB
BPPM	2046	219	89.3	9.7
8-PPM	728	141	93	11.61
256-PPM	308	80	96	14.08

TABLE 2

Average number of photons per ppm bit required achieving 10^{-9} in weak turbulence.

M-ary PPM	Weak turbulence	Weak turbulence with RS code	Improvement due to RS code + multilevel modulation in Weak turbulence respect to non coded BPPM	
			%	dB
BPPM	1462	52	96.4	14.49
8-PPM	543	32	97.8	16.6
256-PPM	258	25	98.3	17.67

All results obtained indicate that, RS codes have low performance in low photons per bit range. This is because the low photons per bit causes burst errors which are beyond the correction capability of RS codes. Under these conditions, spreading the errors by using a code matched interleaver and using concatenated codes will make the coding more effective [12].

4 CONCLUSION

In this paper, the performance of FSO system with M-ary PPM based on RS codes scheme has been numerically analyzed in weak and strong atmospheric turbulence. The results show that, the average number of photons per bit at 10^{-9} BER has been improved by 14.08 dB (compared with the value of BPPM without coding) in strong turbulence and to 17.67 dB in weak turbulence using RS (255,207) + 256-PPM.

The results also show that, coded strong turbulence outperforms the non-coded weak turbulence without coding for BPPM, 8-PPM and 256-PPM at 10^{-9} BER, which indicates a great improvement of the performance of the system tolerance for the intensity fluctuations induced by atmospheric turbulence. RS codes can be combated with matched interleaver concatenated coding to solve the problem in low photons per bit range.

ACKNOWLEDGMENT

Authors wish to thank Prof. Kamran Kiasaleh from University of Texas at Dallas, USA, for his fruitful discussions.

REFERENCES

- [1] Yi Xiang, Liu. Zengji, Yue Peng and Shang Tao, "BER performance analysis for M-ary PPM over gamma-gamma atmospheric turbulence channels," Proceeding of the 6th International Conference on Networking and Mobile Computing (WiCom), Chengdu, pp. 1-4, 2010.
- [2] S. Sheikh Muhammad, Tomaz Javornik, Igor Jelovcan, Zabih Ghassemlooy and Erich Leitgeb, "Comparison of hard-decision and soft-decision channel coded M-ary PPM performance over free space optical links," Eur. Trans. Telecomm., vol. 20, pp. 746-757, 2008.
- [3] S. Hranilovic, Wireless Optical Communication Systems, Springer Science + Business Media, Boston, 2005.
- [4] Zheng Zheng Juanjuan Yan and Anshi Xu Weiwei Hu, "Improved performance of M-ary PPM free-space optical communication systems in atmospheric turbulence due to forward error correction," Proceeding of the 10th International Conference on Communication Technology (ICCT), Guilin, pp.1- 4, 2006.
- [5] Kamran Kiasaleh, "Performance of APD-based, PPM free-space optical communication systems in atmospheric turbulence," IEEE Transactions on Communications, vol. 53, pp.1455-1461, Sep. 2005.
- [6] Mohammed R. Abaza and Kamran Kiasaleh, Erratum to "Performance of APD-based, PPM free-space optical communication systems in atmospheric turbulence," IEEE Transactions on Communications, submitted for publication.
- [7] Philip M. Morse, et al., Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, vol. 10, Milton Abramowitz and Irene A. Stegun, Eds. Washington, D.C.: U.S. Government Printing Office, p. 924, 1972.
- [8] Jagtar Singh and V.K.Jain, "Performance analysis of BPPM and M-ary PPM optical communication system in atmospheric turbulence," IETE Tech. Review, vol.25, no.4, pp. 145-152, July-Aug. 2008.
- [9] J. G. Proakis, Digital Communications, McGraw-Hill, New York, NY, 1983.
- [10] T. Aaron Gulliver, "Matched Q-ary Reed-Solomon Codes with M-ary Modulation," IEEE Transactions on Communications, vol. 45, pp.1349-1353, Nov. 1997.
- [11] Murat Uysal and Jing Li, "BER performance of Coded Free-Space Optical Links over Strong Turbulence Channels," Proceeding of the 59th Vehicular Technology Conference (VTC), Milan, pp. 352-356, 2004.
- [12] S. Sheikh Muhammad, et al., "Reed Solomon coded PPM for Terrestrial FSO Links," Proceeding of the 7th International Conference on Electrical Engineering (ICEE), Lahore, pp. 1-5, 2007.

Electricity Sector Restructuring Experience of Different Countries

Archana Singh, Prof. D.S.Chauhan

Abstract— Electricity Market from economic, regulatory and engineering perspective is a very demanding system to control. There is requirement of provision of cost efficiency, lower impact of environment alongwith maintenance of security of supply for use of competition and regulation in the electricity market. Many countries due to failure of its system for adequately management of electricity companies, followed restructuring for its electricity sector. In various countries, different restructuring models were experimented but in the initial phase restructuring was opposed by the parties favouring existing vertically integrated electricity sector. In the paper, restructuring experience of different countries are outlined.

Index Terms—Deregulation, Wholesale Electricity Market, Forward Markets, Independent system Operator, Power Exchange.

1 INTRODUCTION

Electric utilities have been vertically integrated monopolies that have combined generation, transmission and distribution facilities to serve the needs of the customer in their service territories. The price of electricity was traditionally set by a regulatory process, rather than using market forces, which were designed to recover the cost of producing and delivering electricity to customers as well as the capital cost. Due to this monopolistic service regime, customers had no choice of supplier; and suppliers were not free to pursue outside their designated service territories. The main reason for deregulation in developing countries has been to provide electricity to customers at lower prices, and to open the market for competition by allowing smaller players to have access to the electricity market by reducing the share of large state owned utilities. On the other hand, high growth in demand and irrational tariff policies have been the driving forces for the deregulation in developing countries. Technical and managerial inefficiencies in these countries have made it difficult to sustain generation and transmission expansions and hence many utilities were forced by international funding agencies to restructure their power industries[1].

Electricity markets are having a very important characteristic of its organizational structure which has been accommodated as the most significant change in the industry. Vertically integrated industry structure (a regulated monopoly) as the traditional industry structure was owned and operated as a single organization for distribution, transmission, and generation functions[2]. However, the vertically integrated structure, by virtue of the fact that it is a monopolistic structure, is not amenable to introduction of competition.

Current industry structure primarily requires separate functions of the generation and distribution (or con-

sumption) from transmission as considering different functions associated with selling and buying electric energy. The reason behind separation of transmission which is the means of transporting the tradable commodity and ability to influence the transmission- use through, for example, line ratings, line maintenance schedules and network data would be to avoid very powerful competitive advantage to a participant. Beside this, another important function is system operation which is traditionally viewed as a generation/transmission function. This function has evolved to the Independent System Operator (ISO) in the most electricity markets presently which is responsible for coordinating maintenance schedules and performing security assessment.

The deregulation processes have been started with debate for defending the vertically integrated model from opposition by private and state monopolies [3]. The first was Chile to start effort in 1980s for restructuring its electricity sector. The most discussed deregulation was the British one, with more interest in Norway Model and much attention to actions in United States, especially California State. In South America a major transformation took place throughout the electric power industry from 1980 onwards (chronological progress shown in fig.1).

2 RESTRUCTURING EXPERIENCE OF DIFFERENT COUNTRIES

The electricity sector reform in many developed countries have already undertaken since the 1980s. Initially it was not clear to how to increase efficiency by electricity sector reform. As a matter of fact over various countries, there exists diversity in the wholesale electricity market operation. A transparent, open marketplace would encourage competition among generators and reveal the inefficiencies of the current system to improve the efficiency of the

electricity sector.

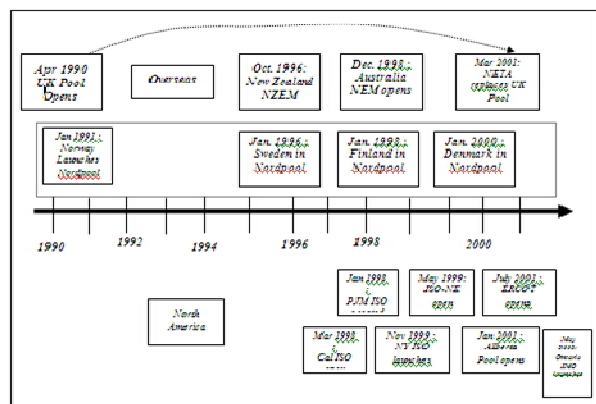


Fig.1. Chronological progression of developments of power markets across the world.

2.1 Chile Market

Hydropower production in Chile varies from 60% to 95%. For supply, there are 4 Gencos with one owned by State Government. Gencos are dispatched depending on audited generation costs and reservoir levels. There are 3 Transcos and private 13 Discos. A poolco which is Economic load dispatch center (CDEC) forecasts global demand and updates values monthly. The system operator in charge of coordinating grid operation is run only by generators favoring the development of several private transmitters. In Chile, wholesale market includes a unified market place, a transport system that "carries" power and market prices defined a spot price for each node along grid. A poolco manages dispatch, reliability and pooling functions. The National Energy Commission (CNE) provides arbitration in the case of disputes [4]. Losses including energy theft were halved in Chile within just seven years and Number of customers per distribution worker has multifolded in ten years after restructuring.

2.2 Colombia Market

Wholesale Electricity Market came into operation in July 1995. Rules promoting free competition in the generation and commercialization business were implemented. Transmission and distribution business were treated as monopolies and competitions were implemented wherever possible. For the reservoirs of the national interconnected system (NIS) defining minimum operative levels a methodology was ruled. CREG (Energy Regulation Commission Group) defined limits for the horizontal and vertical integration business. An energy stock market (generator pool) and a central operator (National Dispatch Centre) of the NIS to support its operation. Transac-

tions between transmission and commercialization bidders are made under two modalities which are through subscription of guaranteed bilateral contracts and through direct transactions in the energy stock market, free offer and demand. Projects of expansion plan are assigned through a scheme of public bids. There is open access to the NTS (National Transmission System) network [5].

Transmission charges are based on connection charges and use of network charges. Evolution of the electricity sector has multifolded by significant increase in private participation. Competition in commercialization to unregulated users has incremented.

2.3 Argentina Market

Power sector restructuring activities in Argentina started in 1990, resulting in the enactment of the Electric Power Regulatory Frame. It created the National Regulatory Agency (ENRE) and wholesale Power Market. Transco and Disco required license to operate. Generation developed as free activity in competition and having the transmission network as open access. Marginal declared costs was basis for dispatch on recognition of remuneration for capacity as a function of system failure risks. Transport is organized as a monopolistic activity with national network (Transener) and six concessionaires for regional network (Distros); it included concession contract, transport tariffs based on the economic cost of losses (node factors) and network unavailability (adaptation factors), plus network O & M costs; expansion at the expense of interested parties; failure penalties as a function of transport charges. It was mandatory for Transport concessionaires to provide nondiscriminatory open access to their transmission system. All existing and future load must to be served by distribution concessionaires in their concession area [6]. Wholesale electricity market is undertaken by private company (Cammesa) whose shareholders are the associations of generators, distributors, transporters, the national state with equal distribution of shares and the large users.

2.4 Australia Market

The process of restructuring of the electricity industry in Australia was initiated in 1991, and by 1998 a National Electricity Market was developed, where the National Electricity Management Company (NEMCO) acted as both the ISO and IMO. Generators could sell energy either by bidding in the spot market, or through formal (bilateral) contracts. The most extensive restructuring is occurring in the South Australia, Victoria, New South Wales and Queensland to form National Electricity Market (NEM). The key aspect of transition process were like elimination of barriers to entry and of barriers to trade between states, creation of pool style (bulk electricity

market) competitive entities in generation and in retail supply and development of regulatory arrangements appropriate to the new regime [7]. Table1 below summarizes the deregulated structure in Australia.

	NSW(New South Wales)	Victoria	South Australia	Queensland
Generation	NSW's three ex-ECNSW(Electricity Commission)generating companies and SMHEA(Snowy mountain hydro electricity authority)	Five ex-SECV(State Electricity Commission) generating companies plus SMHEA	ET-SA(Electricity Transmission System Authority) Generation corporation trading as optima energy.	Three generating companies.
Transmission wires	Transmission company:Trans grid	Transmission company:Power Net Victoria	Transmission company:ETSA Transmission corp.	Transmission company.
Bulk Market	NSW Region of NEM1	Victoria region of NEM1	Participating in the Victorian Region of NEM1.	State poll,separate market company
Distribution wires	6 distributors with ring-fenced retailers	5 distributors with ring-fenced retailers	ETSA power corporation	7 distribution wires business
Retail Supply	6 host retailers and unlimited independent supply licenses	5 host retailers and unlimited supply licenses	ETSA power corp.(host retailers);unlimited supply licenses	3 host retailers and unlimited supply licenses

Table1.Electricity Restructuring in Australia

2.5 Nordic Power Market

In Norway, the electricity reforms were initiated in 1991. In 1993, Nordic power exchange was established as an

independent company. Swedish electricity market unbundled in 1996. Thereafter, a common electricity exchange for Norway and Sweden was established under the name of Nord Pool. In 1998, Finland effectively entered into Nordic Market[8]. Denmark joined Nord Pool subsequently. Nord Pool is owned by the Transmission System operators (TSO) of Norway and Sweden. Nord Pool provides freedom of choice to the large consumers. It organizes trade in standardized physical and financial power contracts. Close cooperation between the system operation and market operation is the key feature of Nord Pool. Major contractual relation among Nordic countries is given in figure 2.

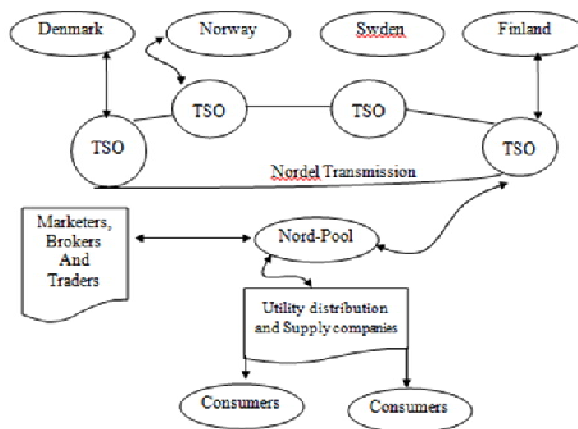


Fig2. Nordic Market – Major Contractual Relationship

2.5.1 Nord Power (Pool Group)

(i) Nord Pool Spot :

It consist of Nord Pool Spot AS and its wholly owned subsidiary Nord Pool Finland Oy, operates the physical day-ahead market Elspot in whole Nordic region and the physical intra-day market Elbas in Finland, Sweden and Zealand (Eastern Denmark). Elspot and Elbas are Nord Pool Spot auction based markets for trade in power contracts for physical delivery. On Elspot, hourly power contracts are traded daily for physical delivery in the next day's 24-hour period. On Elbas, continuous adjustment trading in hourly contracts can be performed until one hour before the delivery hour. Its function is to be the aftermarket to the Elspot market at the Nord Pool.

(ii) Nord Pool ASA - Financial Market :

Nord Pool Financial market is a regulated market place which trade in standardized derivative instruments like forward and future contracts going out several years, and has now started trade in options. Outside this market, there is quite a large and liquid market for over the counter forward and option contracts. The objective of

financial market is to provide an efficient market, with excellent liquidity and a high level of security to offer a number of financial power contracts that can be used profitably by a variety of customer groups. This market is wholly owned by Nord Pool Group.

(iii) Nord Pool Clearing ASA :

It is a licensed and regulated clearing-house. It is central counter party for all derivative contracts traded through exchange and OTC. It guarantees settlement for trade and anonymity for participants. It is wholly owned subsidiary of Nord Pool Group.

(iv) Nord Pool Consulting AS :

It is a consulting firm specializing in development of power market worldwide. It is also a wholly owned subsidiary of Nord Pool Group.

(v) Nordel:

Nordel is an association for electricity cooperation between forum for market participants, nordic system operators and TSOs of nordic countries. The primary objectives of organization are to create and maintain the necessary conditions for an effective nordic electricity market.

2.5.2 Nord pool Features

Nord pool is first multinational commodity exchange for Electric sector in the world. It provide open market to all Nordic Countries with common framework. -Nordic and European markets are example of decentralized day ahead spot market. There are no general cross border tariff among Nordic Countries. Trading of electricity generated by hydropower dominates the cross-border exchanges between the Nordic countries. The balance of electricity trade between the four countries depends on rainfall conditions because of great variation in fuel type capacity of Nordic countries. If hydropower potential is good, Sweden and Norway record trade surpluses, if hydro resource is poor Denmark and Finland will benefit from the electricity trading. -There are only one Market Operator (MO)-Nord Pool and five System Operator (SO) which are Svenska Kraftnät in Sweden, Fingrid in Finland, Statnett in Norway, Eltra in western Denmark, and Elkraft System in eastern Denmark[9]. There are separate regulatory agencies in the four countries. The MO is in principle only responsible for facilitating the trade of electricity as the commodity, but within the physical constraints set by the SO. The operation of the physical system is the sole responsibility of the SO. Further, the market participants

are given the freedom and responsibility of controlling (scheduling) their resources, and have to optimize the utilization of their physical and contractual assets. Transmission system operations are organized on a national basis for Nordic countries. The Five TSOs in the Nordic area are owner of respective main national grid. The National Transmission System Operators (TSOs) are responsible for reliability and balance settlements.

The Elspot market is formed as a day-ahead physical-delivery power market and the deadline for submitting bids for the following day's delivery hours is fixed as 12 am (noon). There are three types of bids available in Elspot; the hourly bid or single bid, block bid and flexible hourly bid. Participants can submit bids to Nord Pool Spot electronically either through EDIEL communication or through the internet application ElwWeb. Nord Pool offers futures contracts for one to nine days ahead and for one to six weeks ahead in time. These futures contracts are settled daily. All these futures and forward contracts use the daily average system price as reference. There are also contracts to hedge zonal price differences, either one quarter or one year ahead. Prices for real-time are determined by the marginal bid like in the day-ahead spot market. Real time market is also known as Regulated Power Market in Norway. Nord Pool PX has a market share of 43% of the physical Nordic demand; the remaining 57% is traded bi-laterally. Nord Pool also operates a trading platform for financial derivatives as well as clearing house for bi-lateral contracts.

Congestion Management is done by market splitting i.e. resolving congestion in day ahead market and counter trade i.e. resolving congestion in real time. Point of Connection tariff structure is followed to promote space. UI pricing mechanism is followed for deviation from schedule. Dr. Per Christer, Senior Vice President, Nord Pool Consultancy has given the Nord Pool Market Model (shown in fig3.).

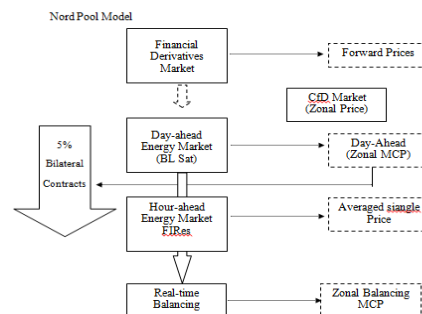


Figure 3. Nord Pool Model as given by Dr. Per Christer

2.6 Pennsylvania – New Jersey – Maryland (PJM) Market

The Pennsylvania – New Jersey – Maryland in-

terconnection (PJM) has been a pool between the three founding utilities that enables co-ordination of trade since 1927. PJM is responsible for the management of a competitive wholesale electricity market across the control areas of its members and for safe and reliable operation of the unified transmission system. All generators defined as a capacity resource in PJM system are obliged to submit an offer into the day-ahead PJM market. Market participants are allowed to self schedule. Transmission system security and reliability considerations are taken into account for the total market clearing operation. A marginal pricing principle is used for market clearing. Each generator at its specific node is paid market clearing price. All loads at their specific nodes are charged as per the market-clearing price. PJM Interconnection is a non-profit company, a limited liability, governed by a board of managers. There is a specific unit 'Market Monitoring Unit (MMU)' within PJM to oversee the functioning of the market. States have public utility commissions (PUCs) and the Federal Energy Regulatory Commission (FERC)(shown in fig4). PUCs regulate generation and distribution's intra-state utility business. The FERC regulates interstate energy transactions including wholesale power transactions on transmission lines[10].

2.7 California State

Public, Political pressure and higher electricity cost have resulted in ending the regulated monopolies of vertically integrated utilities. Deregulation in US proceeded with the Public Utility Regulating Policies Act approval in 1978 and the Energy Policy Act (EPAct) in 1992. Federal Energy Regulatory Commission(FERC) approved non-discriminatory open access to transmission services in 1995. Utilities and Regulators, including American Electric Power (AEP), the California Public Utilities Commission (CPUC), the New England Electric System (NEES) and the Pennsylvania/New Jersey/Maryland(PJM) pool have formulated several proposal for change[11].

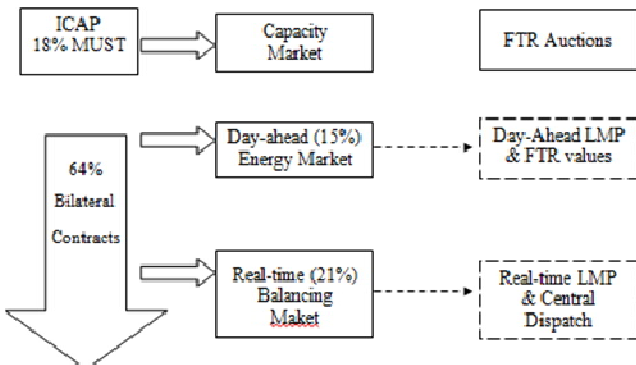


Figure 4. PJM Model

The Comprehensive National Energy Strategy announced on April 1998 and stressed that it relies as much as possible on free markets and competition. An Independent System Operator (ISO) and Power Exchange (PX) have been established in 1998 based on market structure after the CPUC's decision in 1995 which became watershed for the road towards competitive market. The outlining of the proposed California Model filed with FERC on April 29, 1996 with three investor owned utilities (IOUs) in California which were Pacific Gas & Electric or PG&E, Southern California Edison and San Diego Gas Electric. There are three significant characteristic in California Model-

- a) To simplify the transmission pricing scheme including nodal and congestion charge assessing, Zonal Approach is applied.
- b) A Scheduling coordinator (SC) or PX have been introduced to manage multiple separate energy forward markets (each with a supply and demand portfolio). An adjustment approach is adopted to perform inter-zonal congestion management.
- c) An adjustment bid approach is adopted to perform inter-zonal congestion management.

An Independent System Operator (ISO) and a Power Exchange (PX) have been established in 1998 based on market structure and rules governed by FERC. Multiple separate energy forward markets, each with a supply and demand portfolio managed by a Scheduling Coordinator (SC) or PX have been introduced. The total separation of the wholesale power exchange and the market participant was done from ISO.

Power Exchange will be independent entity for managing bid of energy for each half-hour on a day ahead basis for ISO dispatch decision. The ISO will control the power dispatch and the transmission system [12]. It will have no financial interest in the Power Exchange or in any generation, load, and transmission or in distribution facilities. The ISO will coordinate the information exchange in an open market and will work as per North American Reliability Council (NERC) and Western System Coordinating Council (WSSC) reliability standards. The ISO will coordinate day-ahead scheduling and balancing for all users of the transmission grid and also will procure ancillary services.

Scheduling Coordinators (SCs) aggregate participants in the energy trade and are free to use protocols that may differ from pool rules. SCs run a forward market in which parties can bid to buy and sell energy and submit the preferred schedule to the ISO and work with the latter to adjust schedules when necessary (figure 5).

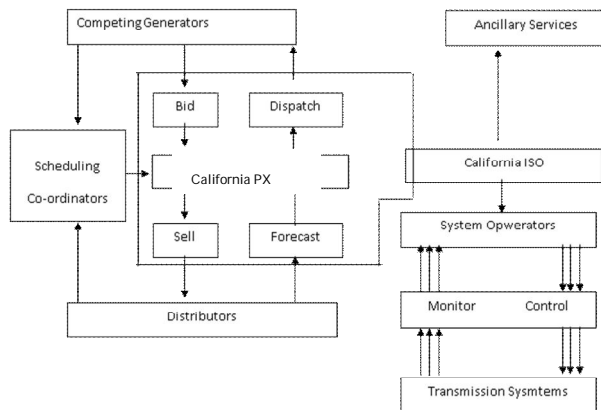


Figure 5. California structure with inclusion of Scheduling Coordinators.

2.8 New Zealand Electricity market

The New Zealand electricity system consists of North and South Island as two alternative current (AC) systems connected by 1200 MW underwater HVDC cable. All capacity on South Island is hydroelectric and it exported power to North Island. Approximately 75% of the North Island demand is met from hydroelectric source with the remaining 25% split between geothermal sources and fossil fuel (coal, gas and oil) sources. Prior to February 1, 1996, the generating industry was dominated by state-owned Electricity Corporation of New Zealand (ECNZ) which owned and operated 95% of total capacity. A wholesale market as Contact Energy Ltd. for electricity was formed as separate state-owned enterprise from ECNZ on February 1, 1996. There are currently 38 electricity distribution companies, providing equal access distribution services and electricity supply to customers and one electricity retailer providing electricity supply only [13]. The wholesale electricity market in New Zealand commenced operation under the name Electricity Market Company (EMCO) on October 1, 1996.

The New Zealand Electricity Market (NZEM) introduces competition within the wholesale electricity sector through creation of a national electricity pool

and a spot market for electricity. The EMCO operates the market through a bidding system and is the clearing house for market transactions. TransPower, the operator and developer of the national grid, performs the various services like provision of reliable national grid, efficient scheduling and dispatch generation to satisfy market demand, purchasing of ancillary services and providing information to the grid users in an open, non-discriminatory manner in the wholesale electricity market.

2.9 Canada market

The restructuring in the Alberta's electric industry was started to retain benefits of the existing low cost generators for customers in 1996. In the new structure, power pool was defined for all energy trade in the province. Generation sector made fully competitive with competitive bidding except the case of old retired power plants. IPP was brought to meet load growth. Grid Company of Alberta (GridCo) administered a province-wide transmission grid. The transmission grid was owned by the four utilities that own transmission facilities in the province and contract all individual owners to supply transmission services. The Electric Transmission Council as advisory group formed to represent the interests of consumers and transmission users [14].

3.0 England & Wales market

UK government took a historic step by privatizing the publicly owned electric power industry in England and Wales (E&W) in 1988. Generation, Transmission and Distribution of electricity were divided into three large companies. All existing fossil fuel plants were taken over by National Power and powerGen.

The National Grid Company (NGC) provides transmission services from generators to the Regional Electricity Supply Companies (RECs) and coordinates transmission and dispatch of electricity generators [15]. NGC runs both the physical and financial side of the E&W electricity market. It serves as both the Independent system operator (ISO) and the Power Exchange and determine both half-hourly market clearing prices and runs the physical national grid, making generator dispatch decisions in real time to manage congestion on the grid and provide ancillary services for reliable supply to all the consumers [16]. NGC uses GOAL (generation ordering and loading)

program to determine the merit order of dispatching generation alongwith reserve capacity[17].

3.1 Indian Market

Indian Power sector is in a transition phase from a regulated sector to a competitive market (taken as author is from India). A competitive market provides the participants with benefits of price determination by market forces, easy access to market, transparent working however it also brings with it many changes that need to be taken care of by the market participants at various stages of development.

The power sector in India has seen significant developments post the enactment of the Electricity Act 2003[18]. The policy and regulatory efforts have also been synchronized to ensure rapid development of the power markets in the country.

In this direction, Electricity Act 2003 has come into force from June 2003 in India. It introduces the concept of trading bulk electricity. The Act has enabled consumers and the distribution companies to have choice in the selection of electricity supplies. Similarly, the generator also has choice to select among the distribution companies (shown in figure 6.). The Act specifies the provisions for non-discriminatory use of transmission lines or distribution system or associated facilities with such lines or system by any licensee or consumer or a person engaged in generation.

At regional level, there are five regional load dispatch centers (NR, WR, ER, SR, NER) which are operated by Power Grid. At state level, there are 28 states which are responsible for their generation, transmission and distribution. States purchase power from Independent State Generation Supply (ISGS). Trade between states is facilitated by trading firms like PTC, NVVL and others[19]. Distribution licenses and Government do not need trading license and transmission licenses and load dispatch centers can not trade power. About 2.5 % of total power generated in country is being traded presently. There are 17 licensed electricity traders for inter state trading till now. Most generation capacity (56% State, 36% CGS, 11% Private) tied up with long term contracts. Only surplus can be traded. The present inter-regional capacity is 11500MW which is planned to go upto 37000 MW upto 2012. India have Pool type centralized mechanism for dispatching central generating plants. On day ahead basis to meet forecast

demand of SEB through respective RLDCs at regional level. But it is mandatory cost based (non bid) Power Pool. ABT mechanism facilitates Balancing market in an inherent way but it has got some limitations also. Current trading occur between ISGS and states STU/SEBS,

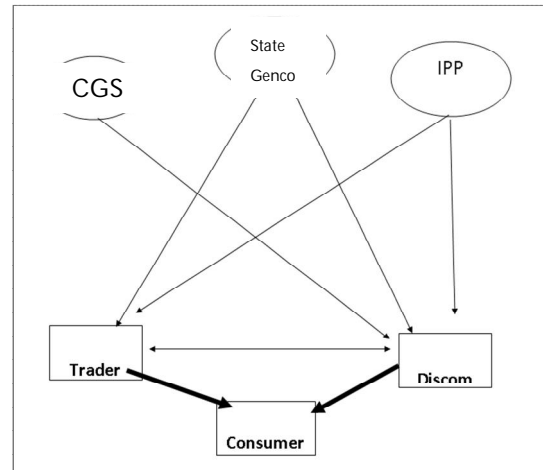


Figure 6. Indian Market Structure after Act-2003

between states, through international import /export (Bhutan, Nepal) and also by state embedded generators/IPPs / Loads and others. The RLDCs organize the day ahead scheduling of the ISGS[20]. Short term bilateral contracts are taking place through traders but they are lacking formal market and real time information. Often Sellers call for separate tenders for surplus available with them and traders compete with each others on prices to get the supply. This situation has resulted in prices of traded power moving only in one direction (higher). The root cause is one-sided competition. On the other hand, buyers are not getting adequate response against tenders called by them. A platform for wide sellers and buyers is not available.

4 CONCLUSION

In the various countries, most of the electric power industry has been going through a process of transition and restructuring by moving away from vertically integrated monopolies and towards more competitive market models since nineties. This has been achieved through as creating competition at each level in the power industry and having a clear separation between its generation, transmission and distribution activities as well. Different countries are im-

plementing industry restructuring in a variety of ways, depending on the characteristics of each market area which include: diversity of generation by fuel types, demand/supply balances, the extent of transmission capacity to facilitate energy imports to meet market demand and etc.. In designing and planning the market structure and rules for competition in their jurisdiction, governments, regulators and other industry participants are influenced by local market characteristics and the practices.

References

- [1] M,Illic, F. Galiana and L. Fink, "Power System Restructuring:Engineering and Economics",Kluwer Academy Publisher,1998.
- [2] <http://class.ece.iastate.edu>.
- [3] A.K.Izaguirre, "Private participation in the electricity sector-recent trends",World Bank report-Private Sector, December 1998,pp.5-12.
- [4] H. Rudnick , "Chile: Pioneer in deregulation of the electric power sector",IEEE Power Engineering Review ,June 1994,pp. 28-30.
- [5] M.I.Dussan , "Restructuring the electric power sector in Colombia",IEEE Power Engineering Review,June 1994,pp21-22.
- [6] C.M. Bastos," Electric energy sector in Argentina",IEEE Power Engineering Review June 1994 ,pp.13-14.
- [7] H. Outhred, "A review of electricity industry restructuring in Australia",Electric Power Systems Research,vol.44,1998,pp.15-25.
- [8] R.D.Christie and I. Wangesteen,"The energy market in Norway and sweden:Introduction",IEEE Power Engineering Review,February 1998,pp.44-45.
- [9] <http://www.nordpool.com>.
- [10] PJM Interconnection LLC,"PJM Open Access Transmission Tariff: Schedule-2", Fourth Revised, Vol1, Issued February 2001.
- [11] Z.Alaywan and J.Alen,"California electric restructuring :a broad description of the development of the California ISO",IEEE Trans on Power Systems ,Vol.13,No.4,November 1998,pp.1445-1451.
- [12] <http://www.caiso.com>.
- [13] T.Alvey,D.Goodwin,X. Ma, D. Streiffert and D Sun,"A security – constrained bid-clearing system for the New Zealand wholesale electricity market",IEEE Trans on Power Systems,Vol.13,No.2,May1998,pp.340-346.
- [14] J.R. Frey,"Restructuring the electric power industry in Alberta ",IEEE Power Engineering Review ,February 1996,pp.8.
- [15] National Grid Electricity Transmission (NGET) plc, "The connection and use of system code (CUSC)," Issued Feb.2006.
- [16] <http://www.nationalgrid.com/uk>.
- [17] <http://www.iitk.ac.in/ime/anoops>.
- [18] <http://www.cerc.org>.
- [19] <http://www.ee.iitb.ac.in/wiki/faculty/sak>.
- [20] <http://www.jbic.go.jp/en/research/report>.

Significant Role of Search Engine in Higher Education

Rahul J. Jadhav, Dr. Om Prakash Gupta, Usharani T. Pawar

Abstract- Information explosion has given a rise to quest for more and more knowledge and its applicability to varied fields. Higher education is not exception to this. The facilities, of varied institutions are perpetually involved in teaching, learning, evaluation and research activities. They are using the modern technology to a great extent. Not only the faculty and academicians but students of today's world also use the latest technology for knowing more and more. Therefore the use of internet has taken rapid stride. For collecting data and information varied programs are developed and the use of search engine prove to be the most significant tool for gathering information and knowledge. Search engine is one of the most widely used method for navigating of cyberspace. The objective of the research paper is to study the significant role of search engine to make the higher education innovative and easily accessible to the students, faculty and researchers.

Keywords: Search engine, Higher education, navigating of cyberspace.

1. INTRODUCTION

Search Engine: A Capsule Description

A **web search engine** is designed to search for information on the World Wide Web and FTP servers. The search results are generally presented in a list of results and are often called *hits*. The information may consist of web pages, images, information and other types of files. Some search engines also mine data available in databases or open directories. Unlike Web directories, which are maintained by human editors, search engines operate algorithmically or are a mixture of algorithmic and human input.

1.1. What is Search Engine?

A program that searches documents for specified keywords and returns a list of the documents where the keywords were found. Although *search engine* is really a general class of programs, the term is often used to specifically describe systems like Google, Alta Vista and Excite that enable users to search for documents on the World Wide Web and USENET newsgroups.

Typically, a search engine works by sending out a *spider* to fetch as many documents as possible. Another program, called an *indexer*, then reads these

documents and creates an index based on the words contained in each document. Each search engine uses a proprietary algorithm to create its indices such that, ideally, only meaningful results are returned for each *query*.

1.2. History of Search Engine

"How could the world beat a path to your door when the path was uncharted, uncatalogued, and could be discovered only serendipitously?" — Paul Gilster, *Digital Literacy*.

History of Search Engine can be said as started in A.D. 1990. The very first tool used for searching on the Internet was Archie (The name stands for "archives" without the "v", not the kid from the comics). It was created in 1990 by Alan Emtage, a student at McGill University in Montreal. The Archie Database was made up of the file directories from hundreds of systems. When you searched this Archie Database on the basis of a file's name, Archie could tell you which directory paths on which systems hold a copy of the file you want. Archie did not index the contents of these sites. This Archie Software, periodically reached out to all known openly available ftp sites, list their files, and build a searchable index. The commands to search Archie were UNIX commands, and it took some knowledge of UNIX to use it to its full capability.

Two other programs, "Veronica" and "Jughead," searched the files stored in Gopher index systems. Veronica (Very Easy Rodent-Oriented Net-

Bharati Vidyapeeth Deemed University, Pune Yashwantrao Mohite Institute of Management, Karad. INDIA E-mail rjjmail@rediffmail.com

Bharati Vidyapeeth Deemed University, Pune Yashwantrao Mohite Institute of Management, Karad. INDIA

Shivaji University, Kolhapur Department of computer science S.G.M College, Karad. INDIA E-mail usharanipawar@rediffmail.com

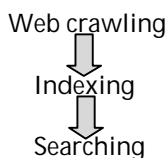
wide Index to Computerized Archives) provided a keyword search of most Gopher menu titles in the entire Gopher listings. Jughead (Jonzy's Universal Gopher Hierarchy Excavation And Display) was a tool for obtaining menu information from various Gopher servers.

In 1993, MIT student Matthew Gray created what is considered the first robot, called *World Wide Web Wanderer*. It was initially used for counting Web servers to measure the size of the Web. The Wanderer ran monthly from 1993 to 1995. Later, it was used to obtain URLs, forming the first database of Web sites called *Wandex*.

In 1993, Martijn Koster created ALIWEB (Archie-Like Indexing of the Web). ALIWEB allowed users to submit their own pages to be indexed. According to Koster, "ALIWEB was a search engine based on automated meta-data collection, for the Web."

1.3. How Search Engine Works?

A search engine operates, in the following order



Web search engines work by storing information about many web pages, which they retrieve from the html itself. These pages are retrieved by a Web crawler (sometimes also known as a spider) — an automated Web browser which follows every link on the site. Exclusions can be made by the use of robots.txt. The contents of each page are then analyzed to determine how it should be indexed (for example, words are extracted from the titles, headings, or special fields called Meta tags). Data about web pages are stored in an index database for use in later queries. A query can be a single word. The purpose of an index is to allow information to be found as quickly as possible. Some search engines, such as Google, store all or part of the source page (referred to as a cache) as well as information about the web pages, whereas others, such as AltaVista, store every word of every page they find. This cached page always holds the actual search text since it is the one that was actually indexed, so it can be very useful when the content of the current page has been updated and the search terms are no longer in it.

This problem might be considered to be a mild form of linkrot, and Google's handling of it increases usability by satisfying user expectations that the search terms will be on the returned webpage. This satisfies the principle of least astonishment since the user normally expects the search terms to be on the returned pages. Increased search relevance makes these cached pages very useful, even beyond the fact that they may contain data that may no longer be available elsewhere.

When a user enters a query into a search engine (typically by using key words), the engine examines its index and provides a listing of best-matching web pages according to its criteria, usually with a short summary containing the document's title and sometimes parts of the text. The index is built from the information stored with the data and the method by which the information is indexed. Unfortunately, there are currently no known public search engines that allow documents to be searched by date. Most search engines support the use of the Boolean operators AND, OR and NOT to further specify the search query. Boolean operators are for literal searches that allow the user to refine and extend the terms of the search. The engine looks for the words or phrases exactly as entered. Some search engines provide an advanced feature called proximity search which allows users to define the distance between keywords. There is also concept-based searching where the research involves using statistical analysis on pages containing the words or phrases you search for. As well, natural language queries allow the user to type a question in the same form one would ask it to a human. A site like this would be ask.com.

2. Importance of Search Engine

The usefulness of a search engine depends on the relevance of the **result set** it gives back. While there may be millions of web pages that include a particular word or phrase, some pages may be more relevant, popular, or authoritative than others. Most search engines employ methods to rank the results to provide the "best" results first. How a search engine decides which pages are the best matches, and what order the results should be shown in, varies widely from one engine to another.

In cyberspace, there's no place to "turn." I have only my computer screen in front of me. Somehow, I need to find a place to purchase the book I want.

There's no street on my screen so I can't drive around on the Web (I could "surf," but that's hit and miss; even then I still need to know where to start). Sometimes it's obvious: type in the name of the bookstore, add a .COM and it's a pretty good bet you're going to end up where you want to go. But what if it's a specialty bookstore and doesn't have a Web site with an obvious URL?

One solution to this problem is the search engine. In fact, it's probably one of the most widely used methods for navigating in cyberspace. Considering the amount of information that's available from a good search engine, it's similar to having the Yellow Pages, a guide book and a road map all-in-one.

Search engines can provide much more information than just the URL of a Web site. Typing in "books" into the Google search engine returns about 9,270,000 results. If we refine the search to "books, Internet", we end up with about 6,070,000 results. If we know the book's author, let's say E.Balguruswamy books, search engines now returns About 80,500 results within 0.18 seconds (of course, these results will change from day to day). For many people, using search engines has become routine. Are the Search Engines are important? Undoubtedly, positively, absolutely....YES! Here's how important they are. In the recent *Georgia Tech Internet User Survey*, respondents were asked how they find pages on the Internet. A full 82% said they used the major search engines.

It is the search engines that finally bring your website to the notice of the prospective customers. When a topic is typed for search, nearly instantly, the search engine will sift through the millions of pages it has indexed about and present you with ones that match your topic. The searched matches are also ranked, so that the most relevant ones come first.

It is the Keywords that play an important role than any expensive online or offline advertising of your website. It is found by surveys that a when customers want to find a website for information or to buy a product or service, they find their site in one of the following ways:

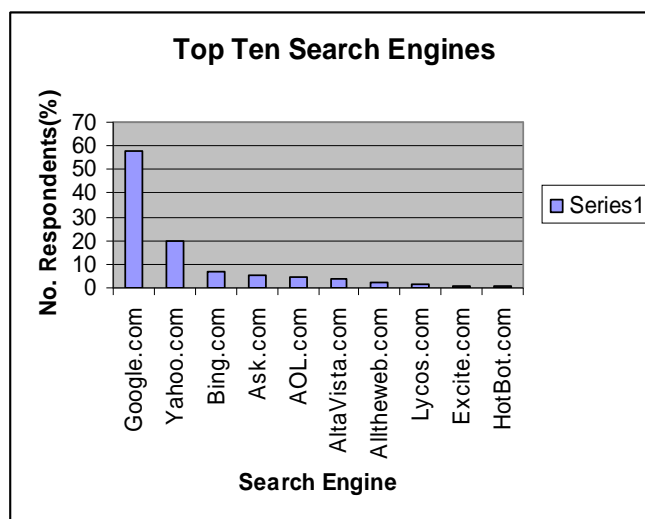
- 1.The first option is they find their site through a search engine.
2. Secondly they find their site by clicking on a link from another website or page that relates to the topic in which they are interested.

3. Occasionally, they find a site by hearing about it from a friend or reading in an article.

Thus it's obvious that the most popular way to find a site is search engine.

Table 1.1 Top Ten Search Engines

Search Engine	No. of Respondents	Percentage
Google.com	112	57.5
Yahoo.com	39	19.5
Bing.com	13	6.5
Ask.com	11	5.5
AOL.com	9	4.5
AltaVista.com	7	3.5
Alltheweb.com	4	2
Lycos.com	3	1.5
Excite.com	1	0.5
HotBot.com	1	0.5



As you can see from the statistics, **Google absolutely dominates the search engine market.** Its closest competitor is Yahoo.com but they seem to be endlessly buying old search technologies that do not provide any innovative techniques. This bodes well for Google's continued dominion.

4. Role of Search Engine in Higher Education

To conduct an effective search, the researcher must understand the structure of the various search engines. Search engines do not always provide the right

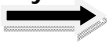
information, but rather often subject the user to a deluge of disjointed irrelevant data.

All search engines support single-word queries. The user simply types in a keyword and presses the search button. Most engines also support multiple-word queries. However, the engines differ as to whether and to what extent they support Boolean operators (such as "and" and "or") and the level of detail supported in the query. More specific queries will enhance the relevance of the user's results.

4.1. Variations on the Search Engine

A search engine is not the same as a "subject directory." A subject directory does not visit the web, at least not by using the programmed, automated tools of a search engine. Websites must be submitted to a staff of trained individuals entrusted with the task of reviewing, classifying, and ranking the sites. Content has been screened for quality, and the sites have been categorized and organized so as to provide the user with the most logical access. Their advantage is that they typically yield a smaller, but more focused, set of results.

Table 2 Search result.

Keyword 	Notes on JAVA	Indian Railway	Bharati Vidyapeeth
Google	56,000,000	2,580,000	272,000
Yahoo	59,600,000	17,200,000	293,000
Bing	8,41,00,000	16,00,000	208,000
AltaVista	92,800,000	17,300,000	64,200
AOL	3,820,000	2,210,000	95,100

Above table 1.2 shows, no search engine covers the entire web. There are technical obstacles such as the inability to index frames, image maps, or dynamically created websites

4.2. Importance of Search in higher education

There can be no downplaying the importance of search in higher education. Search continues to be the number one method for finding relevant information online. Respondents indicated that they spend an average of

almost 4 hours a day online (see table 1.3) and the majority of that time is spent at work (see table1.4)

Table 3 Daily Time spent online

Daily Time spent online	Percent
< 1 hour	10.4
1-2 hours	24.8
2-3 hours	33.1
3+ hours	31.1

Table 4 Work VS Personal Internet Use by the Respondents

Work VS Personal Internet Use By The Respondents	Percent
0% personal/ 100%work	7.8
50% personal/ 50%work	68.6
75% personal/ 25%work	21.2
100% personal/ 0%work	2.4

Respondents were asked the first place they would go online to learn more about the product or service they were considering. Search was the clear winner over manufacturer's sites and information portals, with 66.3 % of respondents. (see Table 1.5).

Table 5 first place to find out educational information

Where would be the first place you would go online to find out educational information	Percent
Search Engine	66.3
Independent Web site	21.6
Educational Portal	8.3
Other	4.8

With the majority of respondents indicating that search plays a major role in their education, we next asked which engine they would use to launch their search. We fully expected Google to be the winner, but we were surprised by how much they dominated their competition (Table 1.6). An amazing 90.9% chose Google as their engine of choice.

Table 6 Search engine chosen by the Respondents

Search engine chosen by the	Percent
------------------------------------	----------------

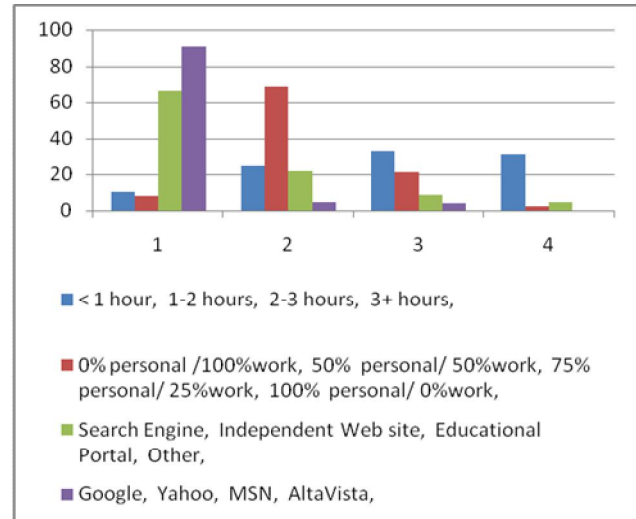
Respondents	
Google	90.9
Yahoo	4.7
MSN	4.3
AltaVista	0.3
Lycos	0.2

5. Conclusion

Table 7 Conclusion

Daily Time spent online			
< 1 hour	1-2 hours	2-3 hours	3+ hours
10.4	24.8	33.1	31.1
Work VS Personal Internet Use by the Respondents			
0% personal /100%work	50% personal/ 50%work	75% personal/ 25%work	100% personal/ 0%work
7.8	68.6	21.2	2.4
First place to find out educational information			
Search Engine	Independent Web site	Educational Portal	Other
66.3	21.6	8.3	4.8
Search engine chosen by the Respondents			
Google	Yahoo	MSN	AltaVista
90.9	4.7	4.3	0.3

This study revealed information that will be key to formulating effective search in higher education. This study also revealed that majority of the respondents search for information on general search engines like Google, yahoo. Google is overwhelming the search engine. Percentage of respondents who search for information relevant to their objective is very low.



6. References

- [1] Craig Lerner. The importance of search engine. Uber Articles March 2, 2010
- [2] Broder, A. Taxonomy of web search, *SIGIR Forum*, vol. 36, pp. 3-10, 2002.
- [3] Web search engine - Wikipedia, the free encyclopedia
- [4] Are Search Engines Important? Searchengineposition.com
- [5] How Do Web Search Engines Work – webopedia 02-17-2006
- [6] LevelTen_Colin The role of Search Engine Optimization in Internet Marketing <http://www.articlesbase.com/>
- [7] The Role of Search in B2B Buying Decisions White paper - Enquiro Search Solutions
- [8] Lee Underwood A Brief History of Search Engines
- [9] Brooks, N. 2004. The Atlas Rank Report II: How searchengine rank impacts conversions. Atlas Institute, 2004.
- [10] Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman G., and Ruppin, E. Placing search in context: the concept revisited. *Proceedings of the WWW Conference*, 2001.
- [11] Greenspan, R. Searching for balance. vol. 2004: *ClickZ stats*.
- [12] R. Villa, M. Chalmers, "A framework for implicitly tracking data", *Proceedings of the Second DELOS Network of Excellence Workshop on Personalisation and Recommender Systems in Digital Libraries, Dublin City University, Ireland, June 2001*.
- [13] iProspect Inc. Search engine user attitudes, 2005.
- [14] Jansen, B., and Spink, A. The effect on click-through of combining sponsored and non-sponsored search engine results in a single listing, *Proceedings of the 2007 Workshop on Sponsored Search Auctions, WWW Conference*, 2007.

Feature Selection for Cancer Classification: A Signal-to-noise Ratio Approach

Debahuti Mishra, Barnali Sahu

Abstract— Cancers are generally caused by abnormalities in the genetic material of the transformed cells. Cancer has a reputation as a deadly disease hence cancer research is intense scientific effort to understand disease. Classification is a machine learning technique used to predict group membership for data instances. There are several classification techniques such as decision tree induction, Bayesian classifier, k -nearest neighbor (k -NN), case-based reasoning, support vector machine (SVM), genetic algorithm etc. Feature selection for classification of cancer data is to discover gene expression profiles of diseased and healthy tissues and use the knowledge to predict the health state of new sample. It is usually impractical to go through all the details of the features before picking up the right features. This paper provides a model for feature selection using signal-to-noise ratio (SNR) ranking. Basically we have proposed two approaches of feature selection. In first approach, the genes of microarray data is clustered by k -means clustering and then SNR ranking is implemented to get top ranked features from each cluster and given to two classifiers for validation such as SVM and k -NN. In the second approach the features (genes) of microarray data set is ranked by implementing only SNR ranking and top scored feature are given to the classifier and validated. We have tested Leukemia data set for the proposed approach and 10fold cross validation method to validate the classifiers. The 10fold validation result of two approaches is compared with hold out validation result and again with results of leave one out cross validation (LOOCV) of different approaches in the literature. From the experimental evaluation we got 99.3% accuracy in first approach for both k -NN and SVM classifiers with five numbers of genes and with 10fold cross validation method. The accuracy result is compared with the accuracy of different methods available in the literature for leukemia data set with LOOCV, where only multiple-filter-multiple wrapper approach gives 100% accuracy in LOOCV with leukemia data set.

Index Terms—Classification, Feature selection, Cancer data, Microarray, Signal-to-noise ratio

1 INTRODUCTION

ALL organisms except viruses consist of cells. Each has one cell, whereas humans have trillions of cells. Each cell consists of a nucleus and inside the nucleus there is DNA, which encodes the programs for making future organisms. Genes make proteins in two steps. First DNA is transcribed to mRNA and mRNA is translated into proteins [1]. Gene expression is the activation of genes that results in a protein. Proteins are the blueprints for the characteristics of the living organisms.

A microarray is a sequence of dots of DNA, protein, or tissue arranged in an array for easy simultaneous analysis. The most famous is the DNA microarray, which plays an integral role in gene expression profiling. The substrate material is glass, plastic or a silicon chip. Important applications of microarrays include the identification of genetic individuality of tissues or organisms, the diagnosis of genetic and infectious disease [2][3].

Debahuti Mishra, ITER, Siksha O Anusandhan University, Odisha, India.
Email: debahuti@iter.ac.in
Barnali Sahu, ITER, Siksha O Anusandhan University, Odisha, India.
Email: sahu.barnali08@gmail.com

Cancers are caused by abnormalities in the genetic materials of the transformed cells. It mostly results from

acquired mutations and epigenetic changes that influence gene expression. A major focus in cancer research is identifying genetic markers. Clinical diagnosis of cancer based on gene expression data has two main targets: first to achieve the correct diagnosis for a cancer patient with a greatest confidence. Second, to identify the gene responsible for a particular type of cancer, this helps in the diagnosis and prognosis of cancer. These objectives imply to develop best classification models which ensure a true classification of a cancer sample with a low risk of misclassification. Many high level data analysis techniques such as clustering and classification algorithms work better with smaller number of genes. This approach usually covers one or more components of microarray data analysis that include dimensionality reduction through a gene subset selection, the construction of new predictive features and model inference [2].

The goal of this paper is to make an intensive study on the techniques available for finding the patterns among the genes or feature selection using SNR ranking and to analyze the result of our two approaches for feature selection which gives significant meaning to classify the genes which are responsible for cancer disease.

This paper is arranged in the following way: introduction to cancer classification data is given in

section 1, section 2 deals with preliminary concept of microarray, classification techniques, SNR ranking, k -means clustering. Section 3 deals with related work on feature selection of cancer data using SNR approach, section 4 deals with the proposed model, section 5 contains experimental evaluation, section 6 explains the validation and comparison of our work and section 7 concludes the paper.

2 PRELIMINARIES

2.1 Microarray

All cells in an organism carry the same genetic information and only a subset of the genes is active (expressed). Analyzing the gene with respect to whether and to what degree they are expressed can help characterize and understand their functions. It can further be analyzed how the activation level of genes changes under different conditions such as for specific diseases [3][4].

Microarray data are generally high dimensional data having large number of genes in comparison to the number of samples or conditions. There are many efficient methods for the analysis of microarray data such as clustering, classification and feature selection.

Feature selection is the preprocessing task for both clustering and classification. Different types of experiment can be done by microarray technology. Microarray technology measures the expression level of genes. That can be used in the diagnosis, through the classification of different types of cancerous genes leading to a cancer type[5]. Basically, genes of microarray data are treated as features, a set of features (genes) give rise to a pattern. If we could get the correct pattern from the data set it is easier to classify an unknown sample based on that pattern.

2.2 Classification Technique Revisited

Our study is mainly based on feature selection and pattern classification for gene expression data related to cancer diagnosis. There are several classification techniques such as SVM, k -NN, neural network, naïve bayesian, decision tree, random forest, top scoring pair.

k -NN: k -NN is the simplest ML technique for classifying objects based on closest training examples in the feature space[6]. It is instance based learning. It gathers all training data and classifiers often via a majority voting, a new data point with respect to the class of its k -nearest neighbor in the given data set. k -NN obtain the neighbors in the given data set. k -NN obtain the neighbors for each data by using Euclidian or Mahalanobis distance between

pairs of data items. The major advantage of k -NN is its simplicity.

Support Vector Machine (SVM): Support vector machines (SVM) is a supervised learning techniques which analyze data and recognize patterns, used for statistical methods and regression analysis[7]. SVM training algorithm builds a model that predicts whether a new sample falls into one category or the other. SVM model is a representation of the samples as points in space, mapped so that the samples of the separate categories are divided by a clear gap that is as wide as possible. New samples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. Support vector machine constructs a hyper plane or a set of hyper planes in a high or infinite dimensional space, which can be used for classification, regression or other tasks.

2.3 k -means clustering Algorithm

Input: k = Number of clusters
 P = A data set containing n features (n number of genes)

1. Select number of cluster k .
 2. Randomly choose k features from the data set as the initial cluster center.
 3. Repeat until the termination criteria fulfilled
 - 3.1 Assign each feature to one of the clusters according to the similarity measure
 - 3.2 Update the cluster means.
 4. until no change in the value of cluster's mean
-

In this approach we have used Euclidean distance as distance measure.

2.4 Signals-to-Noise Ratio

The signal to noise ratio (SNR) test identifies the expression patterns with a maximal difference in mean expression between two groups and minimal variation of expression within each group [8]. In this method genes are first ranked according to their expression levels using SNR test Statistic. The SNR is defined as follows:

$$\text{Signal to noise ratio} = (\mu^1 + \mu^2) (\sigma_1 + \sigma_2) \quad (1)$$

Where μ^1 and μ^2 denote the mean expression values for the sample class 1 and class 2 respectively. σ_1 and σ_2 are the standard deviations for the samples in each class.

3 RELATED WORK

Wai-Ho et al.[9] presents an attribute clustering method, which is able to group genes based on their interdependence to mine meaningful patterns from microarray data. Gene selection methods used are Attribute clustering, t-value, k-means, Biclustering, MRMR, RBF and Classifiers used are C5.0, Neural Networks, Nearest Neighbor, Naive Bayes. Data sets used for the experiment are Colon cancer and Leukemia. Supoj Hengpraprom et al.[10] proposed a method which yields higher accuracy than using the SNR ranking alone and higher than using all of the genes in classification.

Selection of informative features using k-means and SNR ranking. DLBCL, Ovarian, Colon, Prostate, Breast cancer, CNS, Leukemia, Lung Cancer are the data sets used for the experiments. Hualong Yu et al. [11] demonstrated that a modified discrete PSO is a useful tool for selecting marker genes and mining high dimensional data.

SNR ranking is used to select top ranked informative genes. Then PSO is applied to select few marker genes. SVM is used for evaluation of prediction. Colon cancer data set is used for the experiment. Yukee Leung et al.[12] make use of multiple filter and multiple wrappers to improve the accuracy of the classifiers.

Some of MFMW selected genes have been conformed to be biomarkers. Multiple filters are SNR, Pearson correlation, t-statistics. Multiple wrappers are SVM, WV, 3NN and data sets used are LEU [13], COL62 [14], BRER49 [15], LYM77 [16], PROS102 [17], LUNG182 [18]. Shamsul Huda et.al. [19] proposed a hybrid wrapper and filter feature selection algorithm by introducing filters feature ranking score in wrapper stage to get a more compact feature set. They have hybridized mutual information based maximum relevance filter ranking method with artificial neural network based wrapper approach to get the accuracy. Chenn-jung Huang et.al [20] Have under gone a comprehensive study on the capability of probabilistic neural network associated with SNR scoring method for cancer classification. The experimental results show that the combination of PNN with the SNR method can achieve better results for Leukemia data set.

4 PROPOSED MODEL

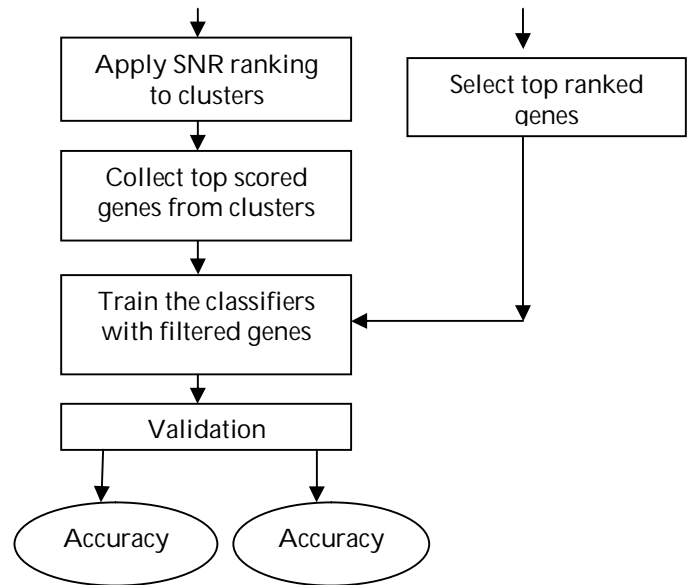
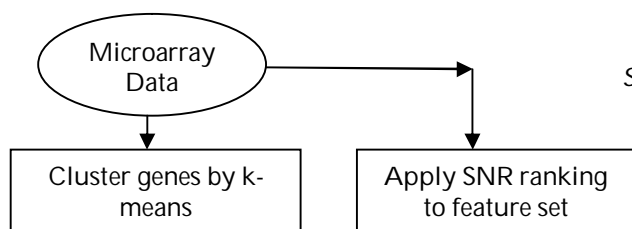


Fig.1. Model for the comparison of accuracies of SVM and kNN in two approaches with 10 fold cross validation method

5 EXPERIMENTAL EVALUATION

We have used leukemia data set of cancer microarray data from Biological data analysis web site [21]. The Data set contains 7,129 genes and 72 samples (47 ALLs, 25 AMLs). For our approach we have taken 50 genes and 72 samples (47 class1, 25 class2) of original data set. The experiment is done in MATLAB version 7.6.0.324 (R2008a), windows XP, PC of Intel Pentium dual CPU. We have implemented two different approaches of feature selection used for classification model to discover differentially expressed genes.

5.1 First Approach for Feature Selection

Step 1: First, the features of data are clustered by applying k-means clustering algorithm. As by applying clustering technique we can group similar type of features in same cluster so that best feature from each cluster can be selected. In our approach we have tested the model with 5, 10 and 20 clusters.

Step 2: The features in each cluster is ranked by applying signal-to-noise ratio scoring technique, so that differentially expressed genes can be easily extracted from each cluster.

Step 3: After that best scored feature in each cluster is selected. We can assure that applying SNR and selecting the best scored feature from each cluster the resultant feature gene subset have no

redundancy.

Step4: The data with the selected biomarkers are tested by different classifiers. The classifiers used are 3NN, SVM

Step5: For validation we have used 10fold cross validation approach. The performance of different classifiers with respect to the number of clusters is given in table 1

TABLE 1
ACCURACY OF SVM AND K-NN IN FIRST METHOD WITH DIFFERENT CLUSTERS

From the above table 1 we can see that both SVM and kNN classifiers are giving same accuracy with 5 numbers of genes in 10fold cross validation method i.e 99.3%. The comparison of accuracy of two classifiers are given bellow in fig.2

Method	Data set	No of clusters	10fold CV accuracy (%)
Kmeans+SNR+ SVM	Leukemia	5	99.3
		10	94.1
		20	96.1
Kmeans+SNR+ kNN	Leukemia	5	99.3
		10	89.3
		20	94.9

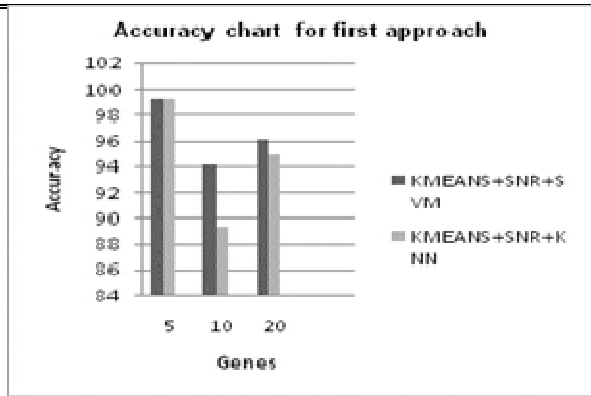


Fig.2. Accuracy of SVM and kNN in first approach with 10fold cross validation

5.2 Second Approach for Feature Selection

Step1: SNR scoring technique is applied to the rows of the data set.

Step2: Basing on the SNR score the features are ranked.

Step3: 5, 10, 20 top scored features are chosen randomly.

Step4: The new data set with these features are fed to

different classifiers independently.

Step5: 10fold cross validation accuracy is listed in table 2.

TABLE 2
ACCURACY OF SVM AND K-NN IN SECOND METHOD AND SVM

Method	Data set	No of genes	10fold CV accuracy (%)
SNR+SVM	Leukemia	5	97.5
		10	96.1
		20	91.4
SNR+ k-NN	Leukemia	5	95.4
		10	90.0
		20	98.1

From table 2 we can observe that kNN classifier with 20 genes give better result is 98.1% accuracy than SVM classifier in 10fold cross validation approach. The accuracy chart of second approach is given bellow in fig. 2

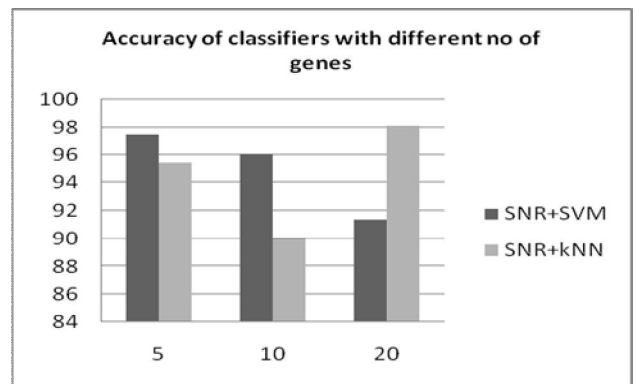


Fig.3. Accuracy of SVM and kNN in second approach with 10fold cross validation

In the paper [22] we have implemented the given two approaches with leukemia data set and validated with k-NN, SVM, PNN and FNN classifiers. We have used hold out validation method to find the accuracy of different classifiers with the above two feature selection approach. In this paper we have considered only two classifiers such as SVM and k-NN the accuracies of these two classifiers with hold out validation method for two approaches are given in table 3 and 4 respectively.

TABLE 3
HOLD OUT VALIDATION ACCURACY OF SVM AND K-NN IN FIRST METHOD

Method	Data set	No of clusters	Hold out validation accuracy (%)
Kmeans+ SNR+ SVM	Leukemia	5	100
		10	96
		20	96
Kmeans+ SNR+ kNN	Leukemia	5	96
		10	83
		20	87

TABLE 4
HOLD OUT VALIDATION ACCURACY OF SVM AND K-NN IN SECOND METHOD

Method	Data set	No of genes	Hold out validation (%)
SNR+SVM	Leukemia	5	96
		10	96
		20	96
SNR+kNN	Leukemia	5	96
		10	96
		20	96

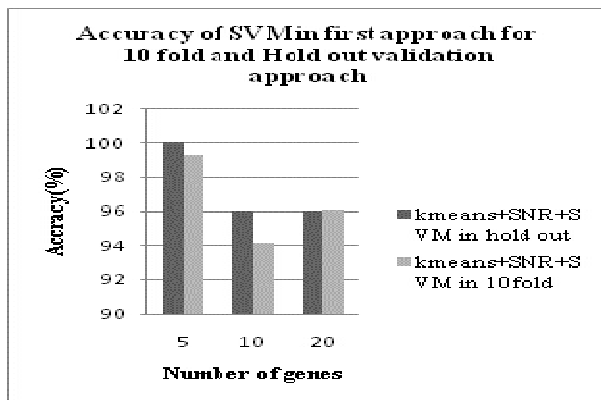


Fig.4. Accuracy of SVM in first approach with hold out and 10fold cross validation

From fig. 4 we can analyze that SVM classifier gives better result in hold out validation method than 10fold cross validation method. But from figure 5 we can conclude that kNN classifier gives better result in first approach with 10fold cross validation method.

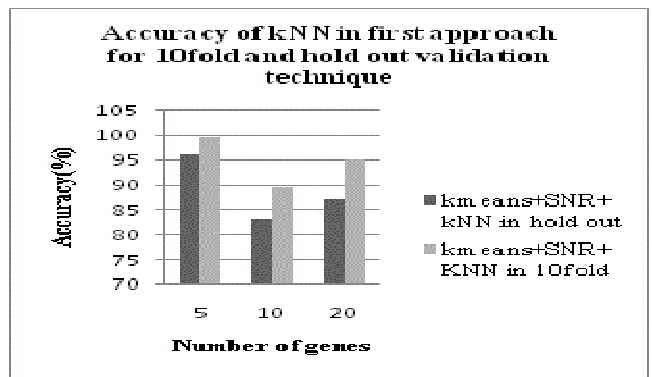


Fig.5. Accuracy of k-NN in first approach with hold out and 10fold cross validation

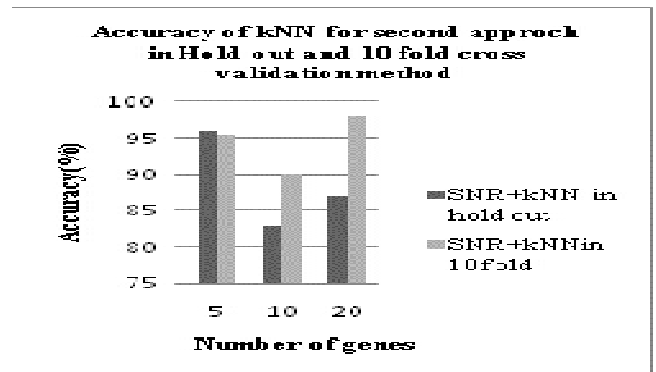


Fig.6. Accuracy of k-NN in second approach with hold out and 10fold cross validation

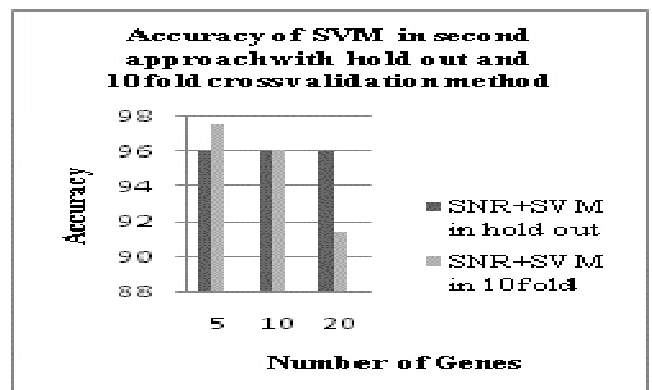


Fig.7. Accuracy of SVM in second approach with hold out and 10fold cross validation

6 COMPARISON AND VALIDATION

Comparing fig.2 and fig.3 we can analyze that the 10fold cross validation accuracy of k-NN and SVM is same i.e 99.3% which is better than the second approach. Where in the second approach SVM gives 97.5% with 5 genes and

k-NN gives 98.1% accuracy with 20 genes .Hence in second approach k-NN gives better result with large number of features

In case of hold out validation method only SVM gives 100% accuracy with and k-NN gives 96% accuracy with 5 numbers of genes.

From the above comparisons we can conclude that SVM gives better results with less number of features in first approach.

Now if we are going to compare the results of two validation method for our two approaches than from fig. 4 and fig.5 we can see that SVM gives better result in hold out validation than 10fold cross validation method for first approach. But k-NN gives better result for 10fold cross validation than hold out validation.

From fig.6 and fig.7 we can analyze that in second approach both k-NN and SVM gives better accuracy in 10fold validation method.

From the literature we have collected the accuracies of different methods or approaches for LOOCV method for Leukemia data set, given in table 5 and the accuracy chart of different approaches are given in fig.8.

Again comparing our approach and methods present in literature we can see that MFMW gives 100% accuracy with 4 numbers of genes and our first approach with Hold out validation with SVM gives 100% accuracy but first approach with 10fold cross validation approach in SVM and k-NN classifier gives 99.3% accuracy with 5 numbers of genes.

TABLE 5

LOOCV ACCURACY OF DIFFERENT APPROACHES FOR LEUKEMIA DATASET

Method	Accuracy(%)LOOCV
MFMW[16]	100
MLP+SNR[11]	76.5
SVM(linear)+SNR[11]	58.8
kNN(Pearson)+SNR[11]	97.1
GPC+ clus[11]	90.3

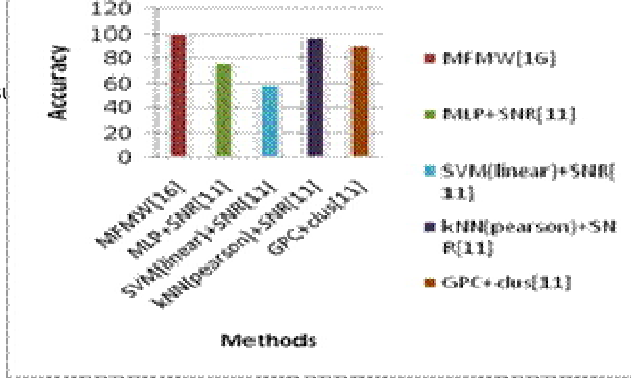


Fig.8. Accuracy of different approaches for Leukemia data set with LOOCV

7 CONCLUSION

From the above comparative analysis we can conclude that our first approach for feature selection is better in comparison to our second approach as due to clustering technique similar features will be grouped in to the same clusters. After applying SNR ranking and selecting top scored features from each cluster may give a true pattern which helps to enhance the classification accuracy. But in case of second approach after applying SNR ranking we can randomly choose the top scored features where we can get redundant feature or noisy features with similar SNR score and does not provide any relevant information about the data. Therefore the performance of the learning algorithms decreases. Again in comparison to hold out validation and 10fold validation approach k-NN and SVM perform well in first approach for 10fold cross validation and only SVM perform well for hold out validation method.

REFERENCES

- [1] Gregory Piatetsky-Shapiro, Pablo Tamayo, "Microarray Data Mining: Facing the Challenges", *SIGKDD Explorations*, Volume5, Issue 2, pp. 1-5, June 2003
- [2] Minca Mramor Gregor Leban, Janez Demar and Bla Zupan, 2007, "Visualization-based cancer microarray data classification analysis", *Bioinformatics*, Vol. 23, No.16, pp.2147-2154, 2007.
- [3] Wolfgang Huber, Anja Von Hey debreck, Martin Vingron, "Analysis of microarray gene expression data, *Hand book of statistics genetics*," 2nd edition, Wiley.2003
- [4] Hong-Hai Do,Toralf Kirsten,Erhard Ralm,"Comparative Evaluation of Microarray-based Gene expression Database,G1-Proceedings, pp 26-34.
- [5] Ana C.loreana, Ivan G.costa, Marcilio c. p. de Souto",*On the complexity of gene expression classification data sets*," Eighth International Conference on Hybrid intelligent System,pp 825-830.2008
- [6] V.N. Vapnik, "Statistical Learning Theory", Wiley-Interscience Publications, 1998

- [7] Vapnik VN. "The nature of statistical Theory". Springer-Verlag:1995
- [8] Miroslava Cuperlovic-Cuf, Nabil Belacel, Rodney. j. Ouellette, "Determination of Tumour marker genes from gene expression data, DDT", Vol-10, Number 6 pp429-437, 2005
- [9] Wai-Ho Au, Keith C.C.Chan, Andrew K.C. Wong, Yang Wang. Attribute clustering for Grouping, IEEE/ACM Transactions on computational biology and Bioinformatics, Vol 2., No 2, pp83-101, 2005
- [10] Supoj Hengprapromh, Prabhas Chongstitvatana, "Selecting Informative Genes from Microarray Data for Cancer Classification with Genetic Programming Classifier using K-Means Clustering and SNR Ranking", *Frontiers in the Convergence of Bioscience and Information Technologies*, pp211-216, 2007.
- [11] Hualong Yu, Guochang Gu, Haibo Liu, Jing Shen, Changming Zhu., "A Novel Discrete Particle Swarm Optimization Algorithm for Microarray Data-based Tumor Marker Gene Selection", *International Conference on Computer science and software Engineering*, pp. 1057-1060, 2008
- [12] Yukyee Leung, Yeungsam Hung, "A Multi-Filter-Multi-Wrapper Approach to Gene Selection and Microarray Date Classification", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vo1. 7, No. 1, pp.108-117, 2010.
- [13] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, and E.S. Lander, "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring," *Science*, vol. 286, no. 5439, pp. 531-537, 1999.
- [14] U. Alon, N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D. Mack, and A.J. Levine, "Broad Patterns of Gene Expression Revealed by Clustering Analysis of Tumor and Normal Colon Tissues Probed by Oligonucleotide Arrays," *Proc. Nat'l Academy of Sciences USA*, vol. 96, no. 12, pp. 6745-6750, 1999
- [15] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J.A. Olson Jr., J.R. Marks, and J.R. Nevins, "Predicting the Clinical Status of Human Breast Cancer by Using Gene Expression Profiles," *Proc. Nat'l Academy of Sciences USA*, vol. 98, no. 20, pp. 11462-11467, 2001.
- [16] M.A. Shipp, K.N. Ross, P. Tamayo, A.P. Weng, J.L. Kutok, R.C.T. Aguiar, M. Gaasenbeek, M. Angelo, M. Reich, G.S. Pinkus, T.S. Ray, M.A. Koval, K.W. Last, A. Norton, T.A. Lister, J. Mesirov, D.S. Neuberg, E.S. Lander, J.C. Aster, and T.R. Golub, "Diffuse Large B-Cell Lymphoma Outcome Prediction by Gene-Expression Profiling and Supervised Machine Learning," *Nature Medicine*, vol. 8, pp. 68-74, 2002.
- [17] D. Singh, P. Febbo, K. Ross, D. Jackson, J. Manola, C. Ladd, P. Tamayo, A. Renshaw, A. D'Amico, and J. Richie, "Gene Expression Correlates of Clinical Prostate Cancer Behavior," *Cancer Cell*, vol. 1, no. 2, pp. 203-209, 2002.
- [18] G.J. Gordon, R.V. Jensen, L.L. Hsiao, S.R. Gullans, J.E. Blumenstock, S. Ramaswamy, W.G. Richards, D.J. Sugarbaker, and R. Bueno, "Translation of Microarray Data into Clinically Relevant Cancer Diagnostic Tests Using Gene Expression Ratios in Lung Cancer and Mesothelioma," *Cancer Research*, vol. 62, no. 17, pp. 4963-4967, 2002.
- [19] Shamsul Hunda, John Yearwood, Andrew Strainieri, "Hybrid wrapper-filter approach for input feature selection using Maximum Relevance and Artificial Neural Network Input Gain Measurement Approximation", *Fourth International conference on Network and system security*, pp442-449, 2010.
- [20] Chenn-Jung Huang, Wei-Chen Liao, "A Comparative Study of Feature Selection Methods for Probabilistic Neural Networks in Cancer Classification", *Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'03)*, Vol 3, pp1082-3409, 2003.
- [21] <http://sdmc.lit.org.sg/GEDatasets/>
- [22] Debahuti Mishra, Barnali Sahu, "A signal to noise classification model for identification of differentially expressed genes from gene expression data," 3rd International conference on electronics computer technology, 2011 (Accepted)

Towards Analysis of Vulnerabilities and Measuring Security Level of Copmputer Network Life Cycle

Irshad Ahmad Mir, Mehraj-U-Din Dar, S.M.K Quadri

Abstract— To measure the overall security of a network, a crucial issue is to correctly compose the measure of individual components. Incorrect compositions may lead to misleading results. The detection of vulnerabilities and security level estimation are the main and important tasks of protecting computer networks. To obtain correct compositions of individual measures, we need to first understand the interplay between network components. For example, how vulnerabilities can be combined by attackers in advancing an intrusion. This paper takes into consideration the models and architecture of intelligent components intended for analyzing computer network vulnerabilities and costing the security level of the computer networks

Index Terms— Security Metrics, Vulnerability Assessment, Attack Model, Security level Assessment.

1 INTRODUCTION

CRUCIAL to today's economy and national security, computer networks plays a central role in most enterprises and critical infrastructures. In protecting these networks against malicious intrusions, a standard way for measuring network security is utmost important. Despite existing efforts in standardizing security metrics [16, 27], a widely accepted network security metrics is largely unavailable. At the research frontier, a qualitative and imprecise view toward the evaluation of network security is still dominant. According to CERT statistic [1] the quantity of attacks on computer networks, their complexity and extent of damage, caused by malefactor's attacks in the Internet, grows each year. The reason is a low security level of majority of systems connected to the Internet. The most common failures exist in operating system (OS) and applications software configuration, software maintenance, user management and administration, including improperly configured OS and applications, incorrect password policy and improper access control settings, existence of vulnerable or easily exploited services and malicious software (Trojans, worms, etc.).

Therefore now vulnerability detection and estimation of security level of computer networks are actual tasks of information assurance. A special class of systems exists for the solution of these tasks vulnerability assessment or security analysis systems (SAS) [17, 20]. The contemporary SAS destine to fulfill checking the system defended against the specified system configuration and security policy for non-compliance and identifying technical vulnerabilities in order to correct them and mitigate any risk posed by these vulnerabilities. The main objective of SAS components is to identify and correct the system management process and security policy failures that produced the vulnerabilities detected. The other important functions are security level estimation, supporting effective interface for control of scanning process, creating reports and automatic updating vulnerability signatures. The SAS components should scan system, update the system configuration according to the specified security policy and system configuration and also send inquiries to modify the security policy if it is necessary. It is also necessary to carry out vulnerability assessment and security analysis during the whole life cycle of computer networks, including initial stage of analysis and design.

In this paper we focus on creating the models, architecture and prototypes of intelligent components of vulnerability detection and security level estimation by expanding the functional capabilities of existing SAS based on simulation and penetration testing. We mainly focus on to design stage. The rest of the paper is structured as follows. Section 2 outlines the approach suggested and related work. Section 3 describes the architecture of security analysis system developed and its implementation issues. Section 4 gives an outline of generalized attack model used for vulnerability assessment and security level estimation. Section 5 describes the model of analyzed com-

- Irshad Ahmad Mir is currently pursuing Ph.D degree program in computer science department in University of Kashmir , India. He did his bachelor's degree in computer application from Amar Singh College Srinagar India and Master degree in computer application from Kashmir university India.
E-mail: irshad.mir@hotmail.com.
- Dr. Mehraj-U-Din Dar is currently the director of Information Technology & Support Systems at University of Kashmir , India.
- Dr. SMK Quadri is Head, PG department of computer sciences, Kashmir University , India. He did his Mtech in computer application from Indian school of Mines and Ph.D in computer sciences from Kashmir University , India.

puter network. Section 6 gives an overview of Experimental approach. Section 7 Draws the conclusion.

2. RELATED WORK

At the design stage, SAS should operate with the model of analyzed computer Network generated from preliminary or detailed design specifications. The main approaches to vulnerability assessment and security analysis can be based on analytic calculation and imitation (simulation) experiments. Analytical approaches use as a rule different risk analysis methods [2, 11, 22, 24, 30, etc.]. Imitational approaches are based on modeling and simulation of network specifications, fault (attack) trees, graph models, etc. [9, 10, 11, 13, 17, 21, 32, 26, 27, 28, 31, etc.].

There are a lot of papers which consider different techniques of attack modeling and simulation: Colored Petri Nets [15], state transition analysis technique [12, 14], simulating intrusions in sequential and parallelized forms [5], cause-effect model [6], conceptual models of computer penetration [29], descriptive models of the network and the attackers [33], structured "tree"-based description [7, 19], modeling survivability of networked systems [18], object-oriented discrete event simulation [3], requires/provides model for computer attacks [32], situation calculus and goal-directed procedure invocation [8], using and building attack graphs for vulnerability analysis [13, 31], etc.

As one can see from our review of relevant works, the field of imitational approaches for vulnerability assessment and security level evaluation has been delivering significant research results. [25] Quantifies vulnerability by mapping known attack scenarios into trees. [16] Suggests a game-theoretic method for analyzing the security of computer networks. The authors view the interactions between an attacker and the administrator as a two-player stochastic game and construct a model for the game. The approach offered in [34] is intended for performing penetration testing of formal models of networked systems for estimating security metrics. The approach consists of constructing formal state/transition models of the networked system. The authors build randomly constructed paths through the state-space of the model and estimate global security related metrics as a function of the observed paths. [31] Analyzes risks to specific network assets and examines the possible consequences of a successful attack. As input, the analysis system requires a database of common attacks, specific network configuration and topology information, and an attacker profile. Using graph methods they identify the attack paths with the highest probability of success. [10] Suggests global metrics which can be used to analyze and proactively manage the effects of complex network faults and attacks, and recover accordingly.

There are a lot of different SAS components which oper-

ate on the stage of exploitation. Examples are NetRecon, bv-Control for Internet Security (HackerShield), Retina, Internet Scanner, CyberCop Scanner, Nessus Security Scanner,

etc. The basic lacks of existing SAS are as follows: (1) use of the scanner does not allow to answer to the main question concerning policy-based systems - "Whether what is revealed during scanning correspond to security policy?"; (2) the quality of obtained result essentially depends on the size and adequacy of vulnerability bases; (3) Implementation of active vulnerability analysis on the computer system functioning in a regular mode can lead to failures in running applications. Therefore not all systems can be tested by active vulnerability analysis.

3. ARCHITECTURE OF SECURITY ANALYSIS SYSTEM

Architecture of security analysis system (SAS) given (fig.1) contains following components: (1) user interface (2) module of malefactor's model realization (3) module of script set (attack scenarios) generation (4) module of scenario execution (5) data and knowledge repository (6) module of data and knowledge repository updating (7) module of security level assessment (8) report generation module (9) network interface.

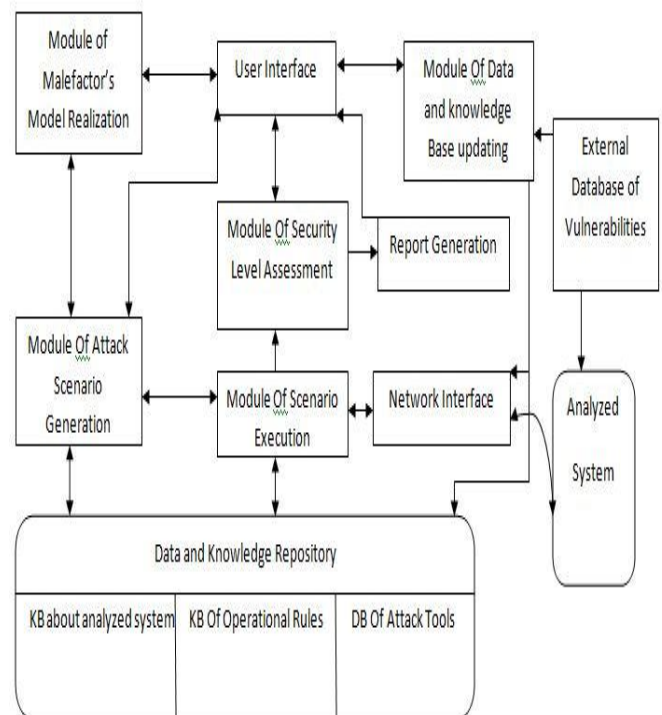


Fig.1. Generalized Architecture SAS

The knowledge base about analyzed system includes data about the architecture and particular parameters of computer network (for example, a type and a version of OS, a list of opened ports, etc) which are needed for scripts generation and attack execution. This data usually can be received by malefactor using reconnaissance actions and methods of social engineering.

The knowledge base of operation (functionality) rules contains Meta and low-level rules of "IF-THEN" type determining SAS operation on different levels of detail. Meta-level rules define attack scenarios on higher levels. Low level rules specify attack actions based on external vulnerability database. IF-part of each rule contains (meta-) action goal and (or) condition parts. The goal is chosen in accordance with a scenario type, an attack intention and a higher level goal (specified in a meta-rule of higher level). The condition is compared with the data from database about analyzed system. THEN-part contains the name of attack action which can be applied and (or) the link on exploit. An example of one of rules is "IF GOAL = "Denial of service" AND OS_TYPE = "Windows_XP" AND OS_VERSION =4 THEN ping_of_death (PoD)". Each rule is marked with an identifier which allows us to determine the achieved malefactor's goal. For example, the rule mentioned above defines a denial of service (DoS) attack "ping_of_death".

The DB of attack tools (exploits) contains exploits and parameters of their execution. A choice of a parameter is determined by the data in KB about analyzed system. For example, the program of ftp brute force password cracking needs to know the ftp server port which can be determined by port scanning.

The module of scriptset (attack scenarios) generation selects the data about analyzed system from the data and knowledge repository, generates attack scriptset based on using operation (functionality) rules, monitors scriptset execution and scriptset updating at runtime, updates data about analyzed system.

The module of scenario execution selects an attack action and exploits, prognoses a possible feedback from analyzed computer network, launches the exploit and recognizes a response of analyzed computer network. In case of interaction with a computer network a real network traffic is generated. In case of operation with the model of analyzed system two levels of attack simulation are provided: (1) at the first level each low-level action is represented by its label describing attack type and (or) used exploit, and also attack parameters; (2) at the second (lower) level each low-level action is specified by corresponding packets of the network, transport and applied level of the Internet protocols stack.

Network interface provides: (1) in case of operation with the model of analyzed system – transferring identifiers and parameters of attacks (or network packets under more detailed modeling and simulation), and also receiving attacks results and system reactions; (2) in case

of interaction with a computer network – transferring, capturing and the preliminary analysis of network traffic. The preliminary analysis includes: (1) parsing of packets according to connections and delivery of information about packets (including data on exposed flags, payload, etc.) and connections; (2) acquisition of data about attack results and system reactions, and also values of some statistics reflecting actions of SAS at the level of network packets and connections.

The module of security level assessment is based on developed taxonomy of security metrics. It is a main module which calculates security metrics based on results of attack actions. The module of database and knowledge repository update downloads the open vulnerability databases (for example, OSVDB - open source vulnerability database) and translates them into KB of operation (functionality) rules of low level

4. GENERALIZED ATTACK MODEL

The model is defined as hierarchical structure that consists of several levels as shown in fig.2. Three higher levels of the attack model correspond to an attacks scriptset, a script and script stages. The scriptset level defines a set of general malefactor's intentions. The second script level defines only one malefactor's intention. Third the set of script stages can contain the following elements: reconnaissance, implantation (initial access to a host), gaining privileges, threat realization, covering tracks and backdoors creation. Lower levels serve for malefactor sub goals refinement. The lowest level describes the malefactor's low level actions directly executing different exploits.

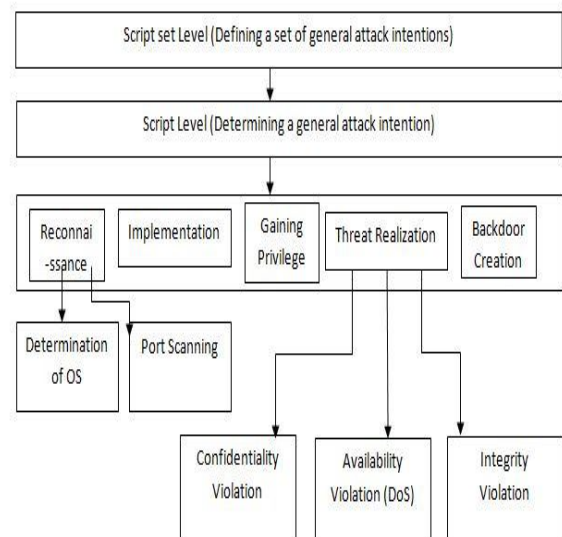


Fig.2. Generalized attack Model

5. ANALYZED NETWORK MODEL

The analyzed computer network model plays a role in evaluating attack results and defining the reaction of the system. It contains following basic components as shown in fig.4 network interface, module of malefactor’s action recognition; module of attack result evaluation; module of system response generation; database of analyzed system; database of attack signatures.

Network interface provides: (1) receiving identifiers and parameters of attacks; (2) transferring attack results and system reactions.

Module of malefactor actions recognition is necessary at realization of detailed attack modeling and simulation, i.e. when malefactor actions are represented as network packets. Functioning of this module is based on a signature method – the data received from the network interface are compared to signatures of attacks from database of attack signatures. Outputs of the module are identifiers and parameters of attacks.

The knowledge base about analyzed system is created from the specification of analyzed system and structurally coincides with KB about analyzed system described in section 2. The difference of these knowledge bases consist in the stored data: KB of the model of analyzed system contains the results of translating the specifications of analyzed system; KB related to the generalized architecture of SAS is initially empty and is filled during the execution of attack scripts.

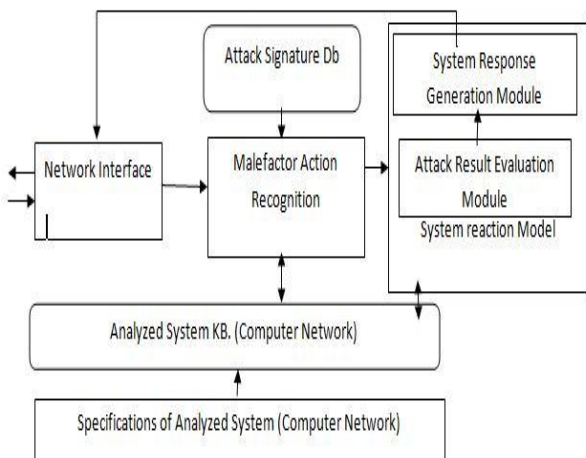


Fig.3. Model of Analyzed Computer Network

5. SECURITY LEVEL ESTIMATION MODEL

The security level evaluation is described by multi-level hierarchy of security metrics . The taxonomy of security metrics is based on the attack model developed. This taxonomy

consists of notions of attack realization action, as well as the notions of type and categories of assets. There are four levels of security metrics sub-taxonomy based on attack realization actions (fig. 4) (1) an integrated level; (2) a script level; (3) a level of the script stages; (4) a level of the threat realization. Each higher level contains all metrics of lower levels (arrow in fig.4 shows the direction of metrics calculation). Examples of security metrics for this taxonomy are as follows: number of total and successful attack scenarios; number of total and successful stages of attack scenarios; number of total and successful malefactor attacks on the certain level of taxonomy hierarchy; number of attacks blocked by existing security facilities; number of discovered and used vulnerabilities; number of successful scenario implementation steps; number of different path of successful scenario implementation, etc.

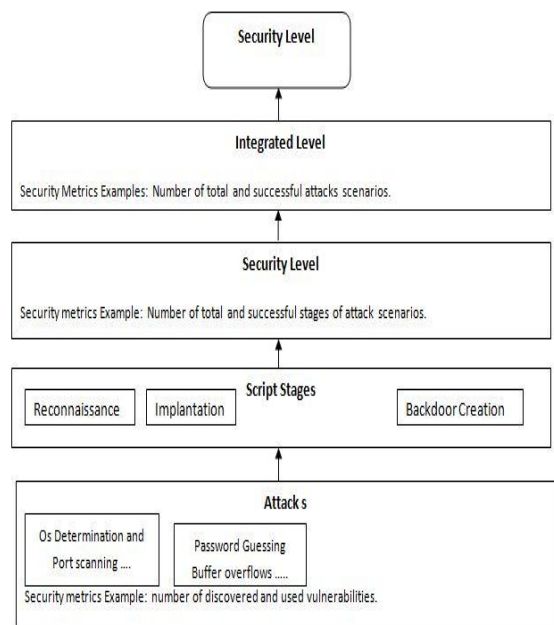


Fig. 4. Security Metrics taxonomy Based on Attack Realization action

6. EXPERIMENTAL APPROACH

We had conducted test based on simulation. The experiment is carried out using “ OPNET MODELER 14.0” which allows us simulate and to form a virtual computer networks. In our test the network consists of three sub-nets :

- Internet area including hosts Int_host and ISP_DNS with ip address 192. 17. 300.*
- A logical sub network including one server with IP address 192.168.0.*.
- Local area network with IP addresses 100.0.0.*.

The basic components (Fig.5) of network are : (1) Internet host with integrated system of software. ;(2) Firewall 1 between logical subnetwork and the Internet ; (3) Mail

server; (4) File server ;(5) Firewall 2 between logical sub-network and LAN ; (6) A Local DNS server, services the clients from LAN; (7) An authentication, , authorization and accounting server; (8) Workstations 1 and 2 . We now consider the SAS prototype for our experiment. We determine the level of the file-server against attacks “ denial of service” taking into consideration that the malefactor’s experience is low . To do this we need to count up the assets and its criticality levels (CRL) and confidentiality levels(COL) . Table.1. shows the CRL and COL of necessary assets.

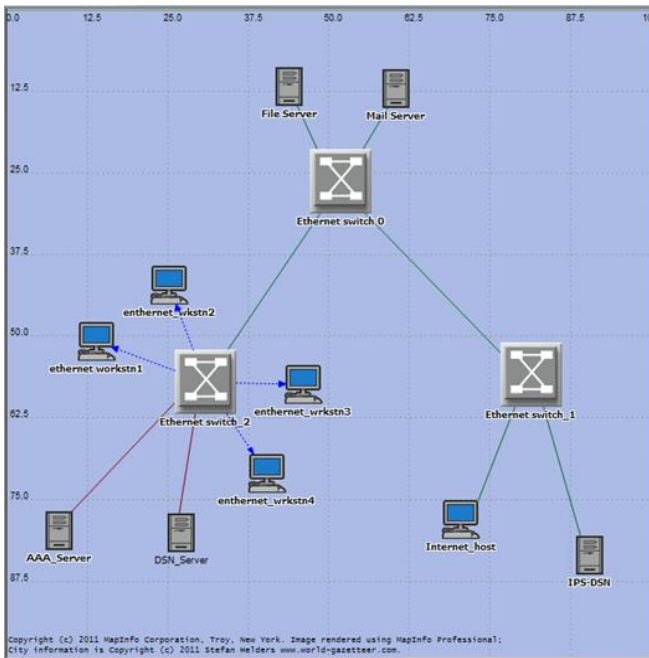


Fig. 5.Configuration of computer Network for Experiment

	Assets	CRL	COL
Information resources:	Information about open ports on file-server	6	3
	Information about file-server OS	6	9
	Information about used service on file-server	6	3
	Information about users on file-server	6	9
	Password of the user “admin”	10	10
Software resources	Operating System	4	9
	Physical resources	0	10
Services	File -server	0	10

Table 1. COR and COL of Assets

According to the malefactor’s model realization SAS creates one script consisting of the following two stages: (1) reconnaissance and (2) threat realization (denial of service). We calculate the confidentiality and criticality levels of successfully attacked assets. At reconnaissance stage, the malefactor has received the information which total level of confidentiality is 10 and total level of criticality is 4. For the information which the malefactor tried to receive the appropriate levels are (20, 20). After normalization, the losses of confidentiality and criticality are (0.6, 0.3). At thread realization stage, the file-server has been successfully attacked (0 points of confidentiality and 10 points of criticality have been lost), therefore the appropriate losses are (0, 1). At script level the losses of confidentiality and criticality are as follows: $((0.6+0)/2, (0.3+1)/2) = (0.3, 0.65)$. The total security metric can be calculated as difference 1 and average value of the given coefficients: $1-0.475=0.525$. Let us select by expert evaluation the following security level scale: (1) “green” – if security level value in an interval [1, 0.8); (2) “yellow” – [0.8, 0.6); (3) “red” – [0.6, 0]. Then the value 0.525 acts as “red” level. As guideline on increase of security level, the report about vulnerability elimination is generated. Procedure of security level evaluation is repeated after eliminating detected vulnerabilities.

7. CONCLUSION

The paper offered the approach to vulnerability analysis and security level assessment of computer networks, intended for implementation at various stages of computer network. This work is based on basic components of intelligent SAS, the model of computer attacks and the model of security level assessment based on developed taxonomy of security metrics. The SAS prototype was implemented and the experiments were held based on simulation of attacks.

REFERENCES

- [1] CERT/CC Statistics 1988-2005. http://www.cert.org/stats/cert_stats.html
- [2] Chapman, C., Ward S.: Project Risk Management: processes, techniques and insights. Chichester, John Wiley (2003)
- [3] Chi, S.-D., Park, J.S., Jung K.-C., Lee J.-S.: Network Security Modeling and Cyber Attack Simulation Methodology. LNCS, Vol.2119 (2001)
- [4]
- [5] Chung, M, Mukherjee, B., Olsson, R.A., Puketza, N.: Simulating Concurrent Intrusions for Testing Intrusion Detection Systems. Proc. of the 18th NISSC (1995)
- [6] Cohen, F.: Simulating Cyber Attacks, Defenses, and Consequences. IEEE Symposium on Security and Privacy, Berkeley, CA (1999)
- [7] Dawkins, J., Campbell, C., Hale, J.: Modeling network attacks: Extending the attack tree paradigm. Workshop on Statistical and Machine Learning Techniques in Computer Intrusion Detection, Johns Hopkins University (2002)
- [8] Goldman R.P.: A Stochastic Model for Intrusions. LNCS, V.2516 (2002)

- [9] Gorodetski, V., Kotenko, I.: Attacks against Computer Network: Formal Grammar-based Framework and Simulation Tool. RAID 2000. LNCS, V.2516 (2002)
- [10] Hariri, S., Qu, G., Dharmagadda, T., Ramkishore, M., Raghavendra C. S.: Impact Analysis of Faults and Attacks in Large-Scale Networks. *IEEE Security & Privacy*, September/ October (2003)
- [11] Henning, R.: Workshop on Information Security System Scoring and Ranking. Williamsburg, VA: *Applied Computer Security Associates and The MITRE Corporation* (2001)
- [12] Iglun, K., Kemmerer, R.A., Porras, P.A.: State Transition Analysis: A Rule-Based Intrusion Detection System. *IEEE Transactions on Software Engineering*, 21(3) (1995).
- [13] Jha, S., Linger, R., Longstaff, T., Wing, J.: Survivability Analysis of Network Specifications. Intern. Conference on Dependable Systems and Networks, *IEEE CS Press* (2000).
- [14] Kemmerer, R.A., Vigna, G.: NetSTAT: A network-based intrusion detection approach. *14th Annual Computer Security Applications Conference*, Scottsdale, Arizona (1998)
- [15] Kumar, S., Spafford, E.H.: An Application of Pattern Matching in Intrusion Detection. Technical Report CSDTR 94 013. Purdue University (1994)
- [16] Lye, K., Wing J.: Game Strategies in Network Security. *International Journal of Information Security*, February (2005)
- [17] National Institute of Standards and Technology. Technology assessment: Methods for measuring the level of Computer security. NIST Special Publication 500-133, 1985.
- [18] Moitra, S.D., Konda, S.L.: A Simulation Model for Managing Survivability of Networked Information Systems, Technical Report CMU/SEI-2000-TR-020, December (2000)
- [19] Moore, A.P., Ellison, R.J., Linger, R.C.: Attack Modeling for Information Security and Survivability. Technical Note CMU/SEI-2001-TN-001. March (2001)
- [20] M. Swanson, N. Bartol, J. Sabato, J. Hash, and L. Graffo. Security metrics guide for information technology systems. NIST Special Publication 800-55, 2003.
- [21] Nicol, D.M., Sanders, W.H., Trivedi, K.S.: Model-Based Evaluation: From Dependability to Security. *IEEE Transactions on Dependable and Secure Computing*. Vol.1, N.1 (2004)
- [22] Peltier, T.R.: Information security risk analysis. Auerbach 2001.
- [23] Peltier, T.R., Peltier, J., Blackley, J.A.: Managing a Network Vulnerability Assessment. Auerbach Publications (2003)
- [24] RiskWatch users manual. <http://www.riskwatch.com>
- [25] Schneier, B.: Attack Trees. *Dr. Dobbs's Journal*, vol. 12 (1999)
- [26] Sheyner, O., Haines, J., Jha, S., Lippmann, R., Wing, J.M.: Automated generation and analysis of attack graphs. *Proc. of the IEEE Symposium on Security and Privacy* (2002)
- [27] Singh, S., Lyons, J., Nicol, and D.M.: Fast Model-based Penetration Testing. *Proceedings of the 2004 Winter Simulation Conference* (2004)
- [28] Steffan, J., Schumacher, M.: Collaborative Attack Modeling. *17th ACM Symposium on Applied Computing (SAC 2002)*, Madrid, Spain (2002)
- [29] Stewart, A.J.: Distributed Metastasis: A Computer Network Penetration Methodology. *Phrack Magazine*, 9 (55) (1999)
- [30] Storms A.: Using vulnerability assessment tools to develop an OCTAVE Risk Profile. SANS Institute. <http://www.sans.org>
- [31] Swiler, L., Phillips, C., Ellis, D., Chakerian, S.: Computer-attack graph generation tool. DISCEX, 2001
- [32] Templeton, S.J., Levitt, K.: A Requires/Provides Model for Computer Attacks. *Proc. Of the New Security Paradigms Workshop* (2000).
- [33] Yuill, J., Wu, F., Settle, J., Gong, F.: Intrusion-detection for incident-response, using a military battlefield-intelligence process. *Computer Networks*, No.34 (2000).

Text Independent Speaker Identification In a Distant Talking Multi-microphone Environment Using Generalized Gaussian Mixture Model

P. Soundarya Mala, Dr. V. Sailaja, Shuaib Akram

Abstract -- In speaker Identification System, the goal is to determine which one of the groups of an unknown voice which best matches with one of the input voices. The field of speaker identification has recently seen significant advancement, but improvements have tended on near field speech, ignoring the more realistic setting of far field instrumented speakers. In this paper, we use far field speech recorded with multi microphones for speaker identification. For this we develop the model for each speaker's speech. In developing the model, it is customary to consider that the voice of the individual speaker is characterized with Generalized Gaussian model. The model parameters are estimated using EM algorithm. Speaker identification is carried by maximizing the likelihood function of the individual speakers. The efficiency of the proposed model is studied through accuracy measure with experimentation of 25 speaker's database. This model performs much better than the existing earlier algorithms in Speaker Identification.

Keywords-- Generalized Gaussian model, EM Algorithm, and Mel Frequency Cepstral Coefficients.

INTRODUCTION

Speaker recognition is the process of recognizing who is speaking on the basis of information extracted from the speech signal. It has been number of applications such as verification of control access permission to corporate database search and voice mail, government lawful intercepts or forensics applications, government corrections, financial services, telecom & call centers, health care, transportation, security, distance learning, entertainment & consumer etc [2].

The growing need for automation in complex work environments and increased need for voice operated services in many commercial areas have motivated for recent efforts in reducing laboratory speech processing algorithms to practice. While many existing systems for speaker identification have demonstrated good performance and achieve high classification accuracy

when close talking microphones are used. In adverse distant-talking environments, however the performance is significantly degraded due to a variety of factors such as the distance between the speaker and microphone, the location of the microphone or the noise source, the direction of the speaker and the quality of the microphone. To deal with these problems micro phone arrays based speaker recognizers have been successfully applied to improve the identification accuracy through speech enhancement [3][4][6].

In speaker identification since there is no identity claim, the system identifies the most likely speaker of the test speech signal. Speaker identification can be further classified into closed-set identification and open-set identification. Speaker identification can be further classified into closed-set identification and open-set identification. The task of identifying a speaker who is known a priori to be a member of the set of N enrolled speakers is known as closed-set speaker Identification. The limitation of this system is that the test speech signal from an unknown speaker will be identified to be one among the N enrolled speakers. Thus there is a risk of false identification. Therefore, closed set mode should be employed in applications where it is surely to be used always by the set of enrolled speakers. On the other hand, speaker identification system which is able to identify the speaker who may be from outside the set of

**Ms P.Soundarya mala Completed M.tech in Digital Electronics and Communication Engineering in GIET in Jawaharlal Nehru Technological University,Kakinada,INDIA in 2010,PH:9493493302.*

Email:soundarya_palivela@yahoo.co.in

**Dr V.Sailaja received Ph.d degree in Speech Processing from Andhra University in 2010, INDIA,PH:9491444434,*

Email:sailajagiet@gmail.com

**Mr Shuaib Akram studying IV B.Tech in Electronics and Communication Engineering, Jawaharlal Nehru Technological University , Kakinada,INDIA,PH:9703976497, Email:shuaib.akram@yahoo.com*

N enrolled speakers is known as open-set speaker identification. In this case, first the closed-set speaker identification system identifies the speaker closest to the test speech data. The speaker identification system is divided into text independent speaker identification and text dependent speaker identification. Among these two, Text Independent Speaker Identification is more complicated in open test.

Speaker Identification:

Given different speech inputs X_1, X_2, \dots, X_c simultaneously recorded through C multiple microphones, whoever has pronounced X_1, X_2, \dots, X_c among registered speakers $S = \{1, 2, \dots, C\}$ is identified by equation (1). Each speaker is modeled by GGMM λ_k .

$$\hat{S} = \arg \max_{1 \leq k \leq C} p(\lambda_k | X_1, X_2, \dots, X_c) \\ = \arg \max_{1 \leq k \leq C} \frac{p(\lambda_k | X_1, X_2, \dots, X_c) \cdot P(\lambda_k)}{p(X_1, X_2, \dots, X_c)} \quad (1)$$

By using Bayes's rule equal prior probability (ie. $P(\lambda_k) = 1/C$), and the conditional independency between different speech inputs X_1, X_2, \dots, X_c given speaker model λ_p , and not in that $p(X_1, X_2, \dots, X_c)$ is the same for all speakers, equation 1 can be simplified as,

$$\hat{S} = \arg \max_{1 \leq k \leq C} \prod_{c=1}^C p(X_c | \lambda_k) \quad (2)$$

Taking the logarithm of equation 2, we obtain

$$\hat{S} = \arg \max_{1 \leq k \leq C} \sum_{c=1}^C p(X_c | \lambda_k) \quad (3)$$

The identity of the speaker can be determined by the sum of hypothesis log likelihood scores obtained from C microphones. In a distance talking environment, however, the log likelihood score itself in equation(3) is expected to degraded, ie. Its reliability cannot be ensured. Furthermore, a variety of causes such as the location of the speaker or the noise, the direction of the speaker, and the distance can have a different effect on each microphone. Therefore the identification result obtained from a microphone can be better than the others. In such cases, the simple integration is greatly affected by the incorrect classification of single channel. Thus, we propose a new

integration method to re-score the hypothesis scores, measure the distance between them and combine them.

2. FINITE MULTIVARIATE GENERALIZED GAUSSIAN MIXTURE SPEAKER MODEL

The Mel frequency cepstral coefficients (MFCC) are used to represent the features for speaker identification. In the set up used, the magnitude spectrum from a short frame is processed using a mel-scale filter bank. The log energy filter outputs are the cosine transformed to produce cepstral coefficients. The process is repeated every frame resulting in a series of feature vectors [1]. We assume that the Mel frequency cepstral coefficients of each are assumed to follow a Finite Multivariate Generalized Gaussian Mixture Distribution. Therefore the entire speech spectra of the each individual speaker can be characterized as a M component Finite multivariate Generalized Gaussian mixture distribution.

The probability density function of the each individual speaker speech spectra is

$$p(\vec{x}_t / \lambda) = \sum_{i=1}^M \alpha_i b_i(\vec{x}_t / \lambda) \quad (4)$$

where, $\vec{x}_t = (x_{tj})_{j=1,2,\dots,D}$; $i=1,2,3,\dots,M$; $t=1,2,3,\dots,T$ is a D dimensional random vector representing the MFCC vector. λ is the parametric set such $\lambda = (\mu, \rho, \Sigma)$ α_i is the component weight such that $\sum_{i=1}^M \alpha_i = 1$

$b_i(\vec{x}_t / \lambda)$ is the probability density of i th acoustic class represented by MFCC vectors of the speech data and the D -dimensional Generalized Gaussian (GG) distribution (M. Bicego et al (2008)) [5] and is of the form

$$b_i(\vec{x}_t | (\mu, \rho, \Sigma)) = \frac{[\det(\Sigma)]^{-1/2}}{[z(\rho) A(\rho, \sigma)]^D} \exp \quad (5)$$

where, $z(\rho) = \frac{2}{\rho} \Gamma\left(\frac{1}{\rho}\right)$ and

$$A(\rho, \sigma) = \sqrt{\frac{\Gamma(1/\rho)}{\Gamma(3/\rho)}} \quad (6)$$

and $\|x\|_\rho = \sum_{i=1}^D |x_i|^\rho$ stands for the

l_ρ norm of vector x , Σ is a symmetric positive definite matrix. The parameter $\vec{\mu}_i$ is the mean vector, the function $A(\rho)$ is a scaling factor which allows the $\text{var}(x) = \sigma^2$ and ρ is the shape parameter when $\rho=1$, the Generalized Gaussian corresponds to a laplacian or double exponential Distribution. When $\rho=2$, the Generalized Gaussian corresponds to a Gaussian distribution. In limiting case $\rho \rightarrow \infty$ Equation (5) Converges to a uniform distribution in $(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)$ and when $\rho \rightarrow 0$, the distribution becomes a degenerate one when $x = \mu$. The

generalized Gaussian distribution is symmetric with respect to μ ,

The variance of the variate x_{ij} is

$$\text{var}(X) = \sigma_{ij} \tag{7}$$

The model can have one covariance matrix per a Generalized Gaussian density of the acoustic class of each speaker. The covariance matrix Σ can also be a full or diagonal. In this chapter the diagonal covariance matrix is used for speaker model. This choice is based on the initial experimental results. As a result of diagonal covariance matrix for the feature vector, the features are independent and the probability density function of the feature vector is

$$b_i(\vec{x}_t | \lambda) = \prod_{j=1}^D \frac{\exp\left(-\left|\frac{(x_{tj}-\mu_{ij})}{A(\rho_{ij}, \sigma_{ij})}\right|^{\rho_{ij}}\right)}{\frac{2}{\rho_{ij}} \Gamma\left(1+\frac{1}{\rho_{ij}}\right) A(\rho_{ij}, \sigma_{ij})} = \prod_{j=1}^D f_{ij}(x_{tj}) \tag{8}$$

The model parameters are estimated and initialized by the EM algorithm with k-means [7].

The updated equation for estimating the model parameters are

The updated equation for estimating α_i is

$$\alpha_i^{(1+1)} = \frac{1}{T} \sum_{t=1}^T \left[\frac{\alpha_i^{(l)} b_i(\vec{x}_t, \lambda^{(l)})}{\sum_{i=1}^M \alpha_i^{(l)} b_i(\vec{x}_t, \lambda^{(l)})} \right] \tag{9}$$

Where $\lambda^{(l)} = (\mu_{ij}^{(l)}, \sigma_{ij}^{(l)})$ are the estimates obtained at the l th iteration.

The updated equation for estimating μ_{ij} is

$$\mu_{ij}^{(1+1)} = \frac{\sum_{t=1}^T t_i(\vec{x}_t, \lambda^{(l)}) A(N, \rho_{ij}) (x_{tj} - \mu_{ij})}{\sum_{t=1}^T t_i(\vec{x}_t, \lambda^{(l)}) A(N, \rho_{ij})} \tag{10}$$

Where, $A(N, \rho_{ij})$ is some function which must be equal to unity for $\rho_i = 2$ and must be equal to $\frac{1}{\rho_{ij}-1}$ for $\rho_i \neq 1$, in the case of $N=2$, we have also observed that $A(N, \rho_{ij})$ must be an increasing function of ρ_{ij} .

The updated equation for estimating σ_{ij} is

$$\sigma_{ij}^{(1+1)} = \left[\frac{\sum_{t=1}^N t_i(\vec{x}_t, \lambda^{(l)}) \left(\frac{\Gamma\left(\frac{3}{\rho_{ij}}\right)}{\rho_{ij} \Gamma\left(\frac{1}{\rho_{ij}}\right)} \right) |x_{tj} - \mu_{ij}^{(l)}|^{\frac{1}{\rho_{ij}}}}{\sum_{t=1}^T t_i(\vec{x}_t, \lambda^{(l)})} \right]^{\frac{1}{\rho_{ij}}} \tag{11}$$

The number of mixture components is initially taken for K – Means algorithm by drawing the histogram of the first Mel frequency cepstral coefficient of the speech data.

3. EXPERIMENTAL SETUP

The proposed Generalized Gaussian Mixture Model in a multi microphone environment are evaluated with a database uttered by 25 speakers. For each speaker, there are 10 conversational sentences, which are recorded in single session. Each sample is of 2 seconds. Speech samples are recorded by using 8 micro phones. Each of which recorded at centre and diagonal. They were then re-recorded again by playing them back with a loud speaker placed at each position, means 1m, 3m and 5m by using 8 microphones which were placed at different locations. These are used to train GGM and to estimate the distribution of average log likelihood estimation.

4. EXPERIMENTAL RESULTS:

As the number of N-best hypotheses per channel (N- best classification results) employed for identification increases, the performance of the proposed method outperform the earlier existing methods.

Table1: Speaker identification accuracy

LOCAL CH	1m		3m		5m	
	C	D	C	D	C	D
0	94.9	95.3	68.3	67.8	75.7	74.6
1	91.8	91.6	66.2	66.9	63.9	64.5
2	94.0	93.4	73.1	72.8	80.6	81.2
3	91.8	92.3	62.7	62.5	56.1	56.9
4	94.6	93.9	54.4	55.6	57.5	58.4
5	94.4	93.8	69.7	70.3	56.5	57.4
6	96.2	95.6	61.8	62.6	60.6	61.4
7	94.0	93.5	57.5	58.9	62.2	63.1
AVG	94.0	93.7	64.2	64.7	64.1	64.7
LS	93.0	93.9	67.4	80.4	70.9	78.3
AD	93.0	93.5	68.3	80.9	74.3	81.3
GGM	94.0	93.7	64.2	64.7	64.1	64.7

The figure represent the identification accuracies of the identification methods, as the number of N- best classification result increases to 25, which corresponds to total number of registered speakers .Table.1 corresponds to the performance of the identification methods in case of N=25. In terms of the base line method (LS), if we approximate k in Eq (3) to k ε N-best hypotheses, it is possible not to obtain the log likelihood score of the hypothesis from a certain channel depending on the order of hypotheses. Thus we considered only when N is equal to 25. Despite only using 2-best classification results per channel, the proposed method (GGM) is comparable to, or even better than, the baseline methods using all possible hypotheses, which consist of all the registered speakers. And the proposed identification method can achieve its best performance with less than eight best hypotheses per channel. Table 2 shows the average percentage of correct identification for various speaker identification models.

Table 2 Avg. percentage of correct identification vs. speaker identification models.

Speaker Model	% of Accuracy
LS	93.0
AD	93.0
GGM	94.0

*LS- Least Squared, AD-Adaptive Distribution,
 GGM- Generalized Gaussian Model

A graph between False Alarm probability and Miss-probability called as DET curve is drawn below. If false alarm probability and Miss-probability are equal, the graph can be approximated as hyperbola.

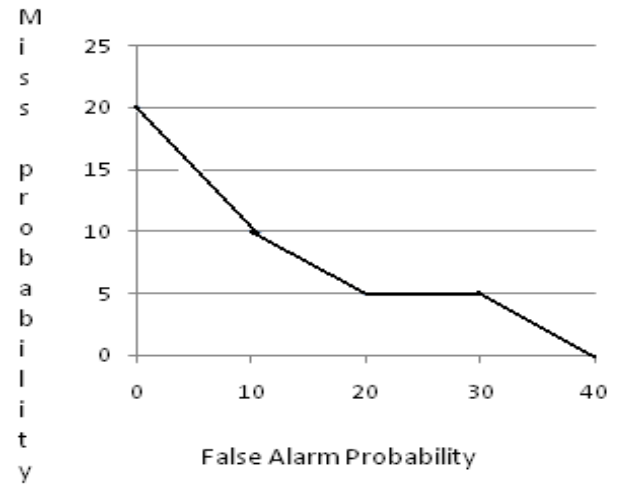


Figure 1: Plot of DET Curves for a speaker identification evaluation.

CONCLUSION

In this paper we propose Text Independent Speaker Identification model based on Finite Multivariate Generalized Gaussian Mixture Model, to improve identification accuracy in a Multi-microphone environment. In a relatively close-talking environment (1m), Generalized Gaussian model maintain high identification accuracy and they are superior to identification result of the best channel. As the distance between the speaker and microphone increases, the Generalized Gaussian Mixture Model shows more reliable identification performance than other existing models which are given in the table (1). For this, a text independent speaker identification model is developed with the assumption that the feature vector associated with speech spectra of each individual speaker follows a Finite Multivariate Generalized Gaussian Mixture Model. The model parameters are estimated and initialized by using EM algorithms with K-means. An experimentation with 25 speakers speech data revealed that this Text Independent Speaker Identification using Finite Multivariate Generalized Gaussian Mixture Model outperform the earlier existing Text Independent models.

References:

- [1] B. Narayana Swamy and R. Gangadharaiah “Extracting Additional Information From GAUSSIAN MIXTURE MODEL PROBABILITIES for improved Text Independent Speaker Identification.” Proc. ICASSP, vol.1, pp.621-624, Philadelphia, USA, March 2005.
- [2] D.A.Reynolds, R.C.Rose. E.M. Hofstetter “Integrated Models of signal and background with application to speaker identification in noise” IEEE Trans. On speech and audio Processing, vol 2. No. 2. April 1994.
- [3] D.A.Reynolds, R.C.Rose. “Robust Text Independent Speaker Identification using Gaussian Mixture Models” IEEE Trans. On speech and signal Processing, 3(1), January 1995, 72-83.
- [4]. I.A. McCowan, J. Pelecanos and S. Sridharan, “Robust speaker recognition using microphone arrays” Proc. Speaker Odyssey, pp.101-106, Crete, Greece, June 2001.
- [5] Md M. Bicego, D Gonzalez, E Grosso and Alba Castro (2008) “Generalized Gaussian distribution for sequential Data Classification” IEEE Trans. 978-1-4244-2175-6.
- [6] Q. Lin, E. Jan, and J. Flanagan, “Microphone arrays and speaker identification”, IEEE Trans. Speech Audio Process, vol.2, no.4, pp.622-628, Oct. 1994
- [7] Sailaja, K Srinivasa Rao & K V V S Reddy (2010) “Text Independent Speaker identification with Multi Variate Gaussian Mixture Model” International Journal of Information Technology and Knowledge Management July-December 2010, Volume 2, No. 2, pp. 475-480.

Novel Object Removal in Video Using Patch Sparsity

B. Vidhya, S. Valarmathy

Abstract— The process of repairing the damaged area or to remove the specific areas in a video is known as video inpainting. To deal with this kind of problems, not only a robust image inpainting algorithm is used, but also a technique of structure generation is used to fill-in the missing parts of a video sequence taken from a static camera. Most of the automatic techniques of video inpainting are computationally intensive and unable to repair large holes. To overcome this problem, inpainting method is extended by incorporating the sparsity of natural image patches in the spatio-temporal domain is proposed in this paper. First, the video is converted into individual image frames. Second, the edges of the object to be removed are identified by the SOBEL edge detection method. Third, the inpainting procedure is performed separately for each time frame of the images. Next, the inpainted image frames are displayed in a sequence, to appear as a inpainted video. For each image frame, the confidence of a patch located at the image structure (e.g., the corner or edge) is measured by the sparseness of its nonzero similarities to the neighboring patches to calculate the patch structure sparsity. The patch with larger structure sparsity is assigned higher priority for further inpainting. The patch to be inpainted is represented by the sparse linear combination of candidate patches. Patch propagation is performed automatically by the algorithm by inwardly propagating the image patches from the source region into the interior of the target region by means of patch by patch. Compared to other methods of inpainting, a better discrimination of texture and structure is obtained by the structure sparsity and also sharp inpainted regions are obtained by the patch sparse representation. This work can be extended to wide areas of applications, including video special effects and restoration and enhancement of damaged videos.

Index Terms— Candidate patches, edge detection, inpainting, linear sparse representation, patch sparsity, patch propagation, texture synthesis.

1 INTRODUCTION

Video inpainting is the process of filling - in of the missing regions in a video. Inpainting technique is the modification of the images in an undetectable form. Nowadays video has become an important media of communication in the world. Image inpainting is performed in the spatial domain where as video inpainting is performed in the spatio-temporal domain.

Video inpainting also plays a vital role in the field of image processing and computer vision as that of the image inpainting. There are numerous goals and applications of the video inpainting technique from the restoration of damaged videos and paintings to the removal or replacement of the selected objects in the video.

Video inpainting is used to remove objects or restore missing or tainted regions present in a video by utilizing spatial and temporal information from neighbouring scenes. The overriding objective is to generate an inpainted area that is merged seamlessly into the video so that visual coherence is maintained throughout and no distortion in the affected area is observable to the human eye when the video is played as a sequence.

Video is considered to be the display of sequence of framed images. Normally twenty five frames per second are considered as a video. Less than twenty five frames per second will not be considered as a video since the display of those will appear as a flash of still image for the human eye.

The main difference between the video and image

inpainting methods using texture synthesis is in the size and characteristics of the region to be inpainted. For texture synthesis, the region can be much larger with the main focus being the filling in of two-dimensional repeating patterns that have some associated stochasticity i.e. textures. In contrast, inpainting algorithms concentrate on the filling in of much smaller regions that are characterised by linear structures such as lines and object contours. Criminisi et al. [3] presented a single algorithm to work with both of these cases provided both textures and structures are present in images.

Nowadays, various researches are performed in the field of video inpainting due to the varied and important applications of an automatic means of video inpainting. The main applications include undesired object removal such as removing unavoidable objects like birds or aeroplane that appear during filming, to censor an obscene gesture or action that is not deemed appropriate for the target audience, but for which it would be infeasible or expensive to re-shoot the scene.

Another main application is the restoration of the videos that are damaged by scratches or dust spots or frames that may be corrupted or missing. When the video is transmitted over unreliable networks, there is a possibility of losing significant portions of video frames. To view the video again in its original form, it is necessary to repair these damaged scenes in a manner that is visually coherent to the viewer.

Initially, all the above applications were performed

manually by the restoration professionals which were painstaking, slow and also very expensive. Therefore automatic video restoration certainly attracted both commercial organizations (such as broadcasters and film studios) and private individuals that wish to edit and maintain the quality of their video collection.

Bertalmio et al. [1], [2] designed frame-by-frame PDEs based video inpainting which laid platform for all researches in the field of video inpainting. Partial Differential Equation (PDE) based methods are mainly edge-continuing methods. In [2], the PDE is applied spatially, and the video inpainting is completed frame-by-frame. The temporal information of the video is not considered in the inpainting process.

Wexler et al [6] proposed a method for space-time completion of large damaged areas in a video sequence. Here the authors performed the inpainting problem by sampling a set of spatial-temporal patches (a set of pixels at frame t) from other frames to fill in the missing data. Global consistency is enforced for all patches surrounding the missing data so as to ensure coherence of all surrounding space time patches. This avoids artefacts such as multiple recovery of the same background object and the production of inconsistent object trajectories. This method provides decent results, however it suffers from a high computational load and requires a long video sequence of similar scenes to increase the probability of correct matches. The results shown are of very low resolution videos, and the inpainted static background was different from one frame to another creating a ghost effect. Significant over-smoothing is observed as well.

Video inpainting meant for repairing damaged video was analysed in [4], [7] which involves gamut of different techniques which made the process very complicated. These works combine motion layer estimation and segmentation with warping and region filling-in. We seek a simpler more fundamental approach to the problem of video inpainting.

Inpainting for stationary background and moving foreground in videos was suggested by Patwardhan et al. [5]. To inpaint the stationary background, a relatively simple spatio-temporal priority scheme was employed where undamaged pixels were copied from frames temporally close to the damaged frame, followed by a spatial filling in step which replaces the damaged region with a best matching patch so as to maintain a consistent background throughout the sequence. Zhang et al., [7] proposed a motion layer based object removal in videos with few illustrations.

In this paper, video inpainting for static camera with a stationary background and moving foreground is considered in the spatial temporal domain. First, the video is converted into image frames. Second, the edges are found by using SOBEL edge detection method. Next, the object to be removed is inpainted using novel exemplar based image inpainting using patch sparsity. The known patch values are propagated into the missing

region for every time frame of image to reproduce the original image. Last, the inpainted image frames are displayed to form the inpainted video. Here a video of short duration is considered for inpainting and the temporal domain information for each image frame is utilized to display the inpainted image frames as a video.

In this paper, Section 2 gives an overview of the image inpainting using extended exemplar-based inpainting method. In Section 3, the method of video inpainting is defined. The experiments and the results are discussed in the Section 4. Finally, conclusion and the future research of the work are discussed in Section 5.

2 IMAGE INPAINTING

The most fundamental method of image inpainting is the diffusion based image inpainting method in which the unknown region is filled by diffusing the pixel values of the image from the known region. Another method of image inpainting is the exemplar based image inpainting in which the region is filled by propagating the patch values of the known region to unknown region. In the previous work of this paper, an exemplar based image inpainting was proposed by incorporating the sparsity of natural image patches.

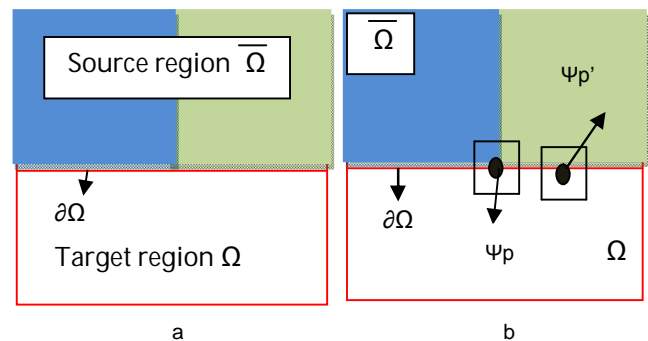


Fig.1 Patch selection

Fig.1 (a) shows that Ω is the missing region, $\bar{\Omega}$ is the known region, and $\partial\Omega$ is the fill-front of patch propagation (b) shows the two examples of the surrounding patch Ψ_p and $\Psi_{p'}$ which are located at edge and flat texture region respectively

The process of filling in of the missing region using the image information from the known region is called as the image inpainting. Let I be the given image with the missing region or target region Ω . In the exemplar based image inpainting, the boundary of the missing region is also called as the fill-front and is denoted by $\partial\Omega$. A patch centered at a pixel p is denoted by Ψ_p .

The exemplar based image inpainting is based on the patch propagation. It is done automatically by the algorithm by inwardly propagating the image patches from the source region into the interior of the target region patch by patch. Patch selection and patch

inpainting are the two basic steps of patch propagation which is to be iterated continuously till the inpainting process is complete.

Fig.1 shows the patch selection process which is used to select the patch with highest priority for further inpainting. The sparseness of the nonzero similarities of a patch to its neighboring patches is called as the structure sparsity which is used for assigning patch priority. As shown in the Fig. 1(b), ψ_p and $\psi_{p'}$ are the patches which are centered at pixel p and p' which lie in the edge structure and the flat texture region of the image respectively. The patch ψ_p has sparser nonzero similarities than patch $\psi_{p'}$, so it is given larger patch priority. The patch inpainting step is performed.

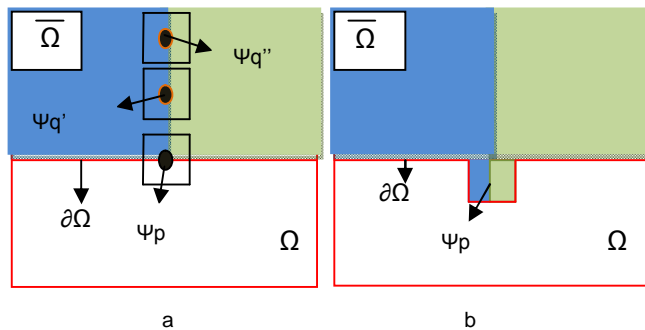


Fig.2 Patch inpainting

Fig.2 (a) shows that for the selected patch ψ_p , sparse linear combination of candidate patches $\{\psi_{q'}, \psi_{q''} \dots \psi_{qN}\}$ is used to infer the missing pixels in patch ψ_p . (b) Shows the best matching patch in the candidates set has been copied into the position occupied by ψ_p , thus achieving partial filling of Ω .

Fig.2 shows the procedure of patch inpainting which is used to select the patch on the boundary for inpainting. The selected patch on the fill-front is the sparse linear combination of the patches in the source region regularized by sparseness prior. In this paper, a single best match exemplar or a certain number of exemplars in the known region to infer the missing patch is not used.

The most likely candidate matches for ψ_p lie along the boundary between the two textures in the source region, e.g., $\psi_{q'}$ and $\psi_{q''}$. The best matching patch in the candidates set has been copied into the position occupied by ψ_p , thus achieving partial filling of Ω . The target region Ω has, now, shrunk and its front has assumed a different shape. Thus image inpainting is performed by the sparse linear combination of candidate patches weighted by coefficients, in which only very sparse nonzero elements exist.

3 VIDEO INPAINTING METHOD

Video is the display of the image frames in sequence. Image inpainting is done in spatial domain, whereas the

video inpainting is performed in both spatial and temporal domain. It is used to remove objects or restore missing regions in the video sequence. Video inpainting may also be considered as the combination of frame by frame image inpainting.

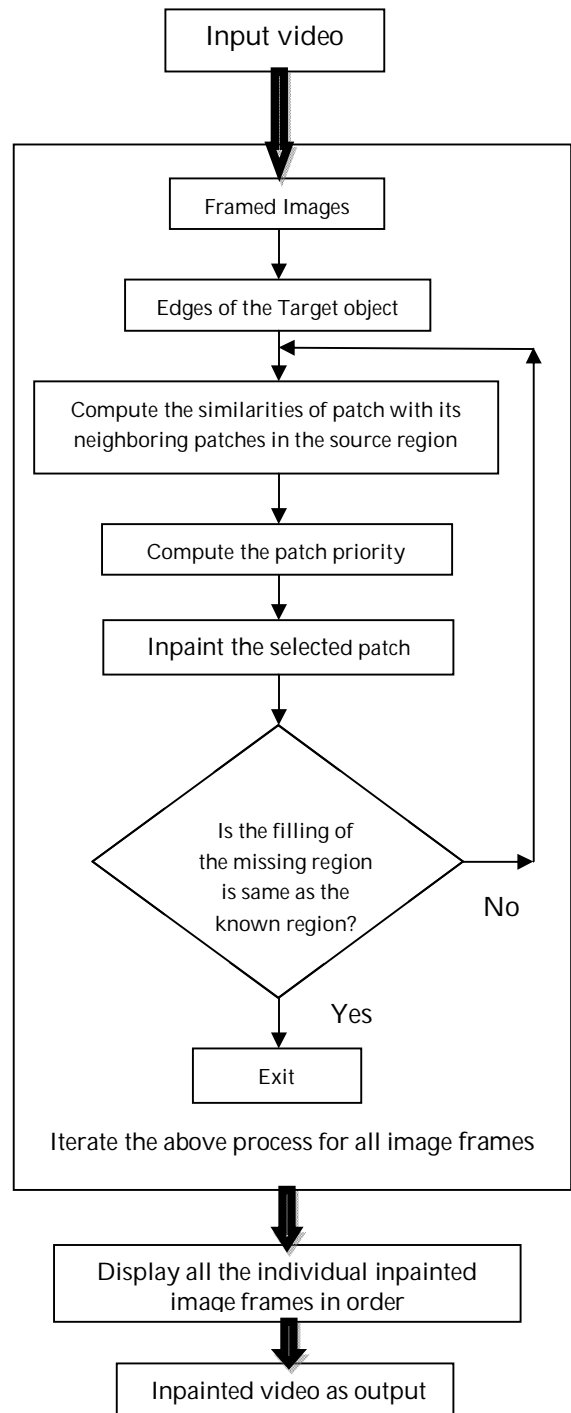


Fig. 3 Flow Diagram of the Video Inpainting

Generally all the natural images are composed of structures and textures. The primal sketches of an image

like the edges, corners, etc are referred to as the structure of an image and the image regions with feature statistics or homogenous patterns including the flat patterns are referred to as the texture of an image.

Patch priority must be defined in such a way that it should be able to differentiate the structures and textures of an image. The structure sparsity is used for assigning priority to the patches to be inpainted. It can also be defined to measure the confidence of a patch located at structure instead of texture.

The structures are sparsely distributed in the image domain. The neighboring patches of a particular patch with larger similarities may also be distributed in the same structure or texture as the patch of interest to be inpainted. To overcome such cases, the confidence of the structure can be modelled for a patch by measuring the sparseness of its nonzero similarities to the neighboring patches.

The input to the object removal algorithm is the framed images. Initially, the structure and texture values of the given image are calculated and the object to be removed is found. Edges of the object to be removed are obtained through the edge detection and the similarities of the patch with its neighboring patches in the source region are computed.

Then, the patch priority is obtained by multiplying the transformed structure sparsity term with patch confidence term. The patch with highest priority is selected for further inpainting. The above process is repeated until the missing region is completely filled by the known values of the neighboring patches.

The sequence of display of the individual time frame of inpainted image is the video inpainting. Image inpainting is performed for each image frame, and all the inpainted image frames are added to form the video. Fig. 3 illustrates the overall flow diagram of the video inpainting process.

4 RESULTS AND DISCUSSION

Fig 4 illustrates the important terms used in the inpainting algorithm. Fig 4. (a) shows the original image I which is given as input to the novel exemplar-based image inpainting method. The natural images are generally composed of structures and textures.

A good definition of patch priority should be able to better distinguish the structures and textures, and also be robust to the orientation of the fill-front. In this paper, a novel definition of patch priority is proposed to meet these requirements. The missing region or the target region is first analysed using the algorithm as shown in the Fig 4. (b).

The structure sparsity is defined to measure the confidence of a patch located at structure instead of texture. For a certain patch, its neighboring patches with larger similarities are also distributed in the same structure or texture as the patch of interest. The patch

which has more sparsely distributed nonzero similarities is prone to be located at structure due to the high sparseness of structures. Confidence of structure for the patch and the data term of the structure are illustrated in the Fig 4. (c) and (d) respectively.

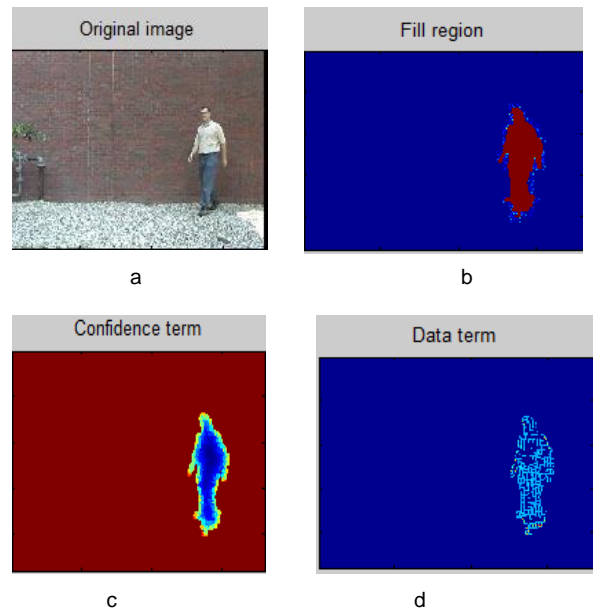


Fig.4 Important terms in the object removal method

Then the patch priority is computed by the product of the data term and the confidence term. The patch Ψ_p with the highest patch priority on the fill-front is selected.

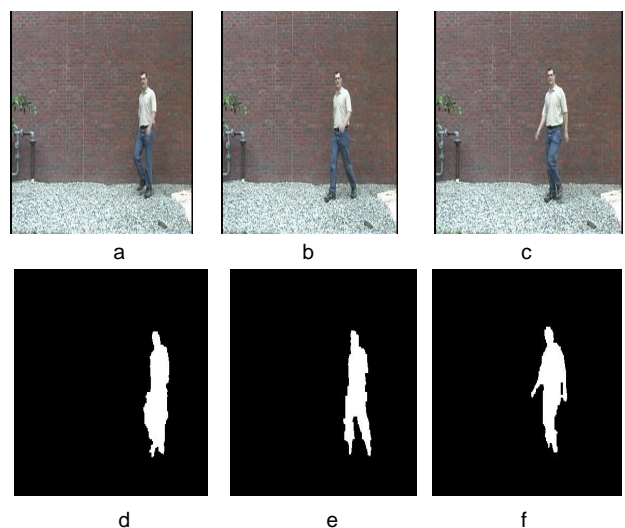


Fig.5 Steps in Video Inpainting

Fig. 5.(a),(b) and (c) shows randomly selected image frame of the original video sequence. Fig 5 (d), (e) and (f) shows edge detection of the image frames (a), (b) and (c) respectively.

Then inpainting is performed based on the patch

propagation by inwardly propagating the image patches from the source region into the interior of the target region patch by patch. The above procedure is repeated for each image frames and displayed in the sequence as video.

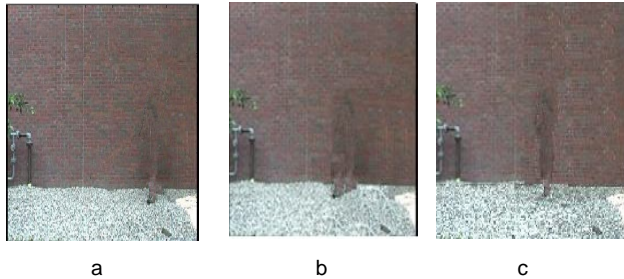


Fig. 6 Inpainted Image frames of the video sequence

Fig 6 (a), (b) and (c) illustrates the inpainted image frames for the corresponding Image Frames shown in Fig 5 (a), (b) and (c) respectively for the given video sequence.

Fig. 5 (a), (b) and (c) illustrates the randomly selected image frames in a video. Fig. 5 (d), (e) and (f) shows the edge detected image frames of Fig. 5 (a), (b) and (c) respectively. The selected patch Ψ_p on the edge is inpainted by the corresponding pixels in the sparse linear combination of exemplars to infer the patch in a frame work of sparse representation. The fill-front $\partial\Omega$ and the missing region Ω are updated instantly. For each newly-apparent pixel on the fill-front, its patch similarities are computed with the neighboring patches and its patch priority.

The selected patch is used for inpainting. By iterating the above process, the filling region is inpainted successfully for the entire image frames. Fig. 6(a), (b) and (c) illustrates the inpainted image frames of the original image frames of the video as shown in Fig. 5 (a), (b) and (c) respectively. The inpainted image frames are added together to display as a video sequence.

5 CONCLUSION

Novel patch propagation based inpainting algorithm for video inpainting is proposed in this paper. It is mainly focussed on the object removal in the image frames. Patch priority and patch representation are the two major steps involved in the proposed exemplar-based inpainting algorithm for an image. Structure sparsity was represented by the sparseness of the patch similarities in the local neighborhood. The patch at the structure with larger structure sparsity is given higher priority and is used for further inpainting. The sparsest linear combination of candidate patches under the local consistency was synthesized by the patch sparse representation. Video is represented by the display of the sequence of the image frames. Hence the inpainted image

frames of each time frame is displayed as the inpainted video. Experiments showed that the proposed exemplar-based patch propagation algorithm can produce sharp inpainting results consistent with the surrounding textures.

In this paper, static camera with constant background is considered. In the future, background with multiple scales and orientations, moving camera and high resolution videos will also be investigated. Video inpainting for long run (time duration) video shall be accomplished. This work can be extended to wide areas of applications, including video special effects and restoration and enhancement of damaged videos.

REFERENCES

- [1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. SIGGRAPH*, 2000, pp. 417-424.
- [2] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-Stokes, fluid dynamics, and image and video inpainting," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, pp. 417-424, 2001.
- [3] Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. Image Process.*, vol. 13, pp. 1200-1212, 2004.
- [4] J. Jia and C. K. Tang, "Image repairing: Robust image synthesis by adaptive and tensor voting," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, pp. 643-650, 2003.
- [5] K.A. Patwardhan, G. Sapiro, and M. Bertalmio, "Video inpainting of occluding and occluded objects", in *Proc. ICIP 2005*. Vol. II, pp. 69-72.
- [6] Y. Wexler, E. Shechtman, and M. Irani, "Space-time video completion," *Proceedings. 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2004.
- [7] Y. Zhang, J. Xiao, and M. Shah, "Motion layer based object removal in videos," *2005 Workshop on Applications of Computer Vision*, 2005.

New Formula of Nuclear Force

Md. Kamal Uddin

Abstract -It is well established that the forces between nucleons are transmitted by meson. The quantitative explanation of nuclear forces in terms of meson theory was extremely tentative & incomplete but this theory supplies a valuable point of view. It is fairly certain now that the nucleons within nuclear matter are in a state made rather different from their free condition by the proximity of other nucleons charge independence of nuclear forces demand the existence of neutral meson as amongst the same type of nucleolus (P-P) or (N-N). This force demands the same spin & orbital angular momentum. The exchange interaction is produced by only a neutral meson. The involving mesons without electric charge, that it gives exchange forces between proton & Neutron & also therefore maintains charge independence character. It is evident for the nature of the products that neutral mesons decay by strong & weak interaction both. It means that neutral mesons constituents responsible for the electromagnetic interaction. Dramatically neutral mesons play an important role for electromagnetic & nuclear force both.

Index Terms - Rest mass energy, Mesons, photons, protons, neutrons, velocity of light, Differentiation

1. INTRODUCTION

It is well established that the forces between nucleons are transmitted by meson. The quantitative explanation of nuclear forces in terms of meson theory was extremely tentative & incomplete, but this theory supplies a valuable point of view. Yukawa first pointed out that nuclear force can be explained by assuming that particle of mass about 200 times the electron mass (mesons) exist & can be emitted & absorbed by nuclear particles (neutrons & protons) with such an assumption a force between nuclear particles of right range & right shape (rapid decrease at large distances is now obtaining).

Now we have the rest mass energy = $m_0 c^2$
Differentiating with respect to r (Inner radius at which nuclear force comes into play)

$$\frac{d(m_0 c^2)}{dr} = \frac{c^2 dm_0}{dr} + m_0 \frac{d(c^2)}{dr} = \frac{c^2 dm_0}{dr} + m_0 \frac{d(c^2)}{dc} \frac{dc}{dr} = \frac{c^2 dm_0}{dr} + 2m_0 c \frac{dc}{dr}$$

. This force is short range, attractive & along the line joining the two particles (central force). (The wide success of this first

application of quantum mechanics to nuclear phenomena gives us confidence in general use of quantum mechanics for the description of the force between heavy particles in nuclei.

Where $dm_0 c^2$ = either rest mass energy of π^0 mesons (For neutral theory), or rest mass energy of π^+ , π^- & π^0 mesons (for symmetrical theory)

dm_0 = either mass of π^0 mesons or mass of π^+ , π^- & π^0 mesons

m_0 = mass of nucleons

$m_0 c^2$ = rest mass energy of nucleons

dr = Range of nuclear force, which can be calculated from differentiation of Nuclear radius. (The force between two nucleons is attractive for distance r (radius) greater than dr (range) & is repulsive otherwise). This strongly suggests & well proved that to some degree of approximation the total isotopic spin T is a constant of the motion & is conserved in all processes, at least with a high probability.

dc = The average velocity of neutron & proton. A large velocity is used in nuclear disintegration.

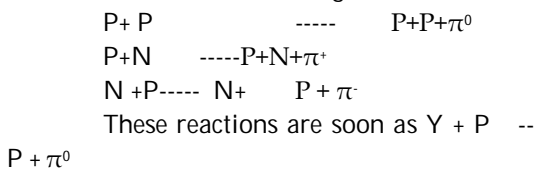
c = Velocity of light

2 = multiplicity of interacting particles is given by (2T+1), the isotopic spin has no such meaning for leptons or a gamma rays

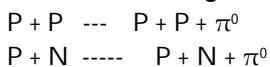
1 = either multiplicity of π^0 mesons or π^+ , π^- & π^0 mesons (evidence of involving of mesons (all type))

Where T = Vector sum of isotopic spin of proton & proton, neutron & neutron, neutron & proton The success of these applications supplies additional support for the hypothesis of the charge independence & charge symmetry of nuclear force. As the nuclear interactions do not extend to very large distances beyond the nuclear radius & this character is useful to solve the problem. The Full charge independence for any system in which the number of neutrons equals the numbers of protons, this formula give the evidence the charge symmetry, merely means that the neutron-neutron & proton proton interaction are equal but says nothing about the relations of neutron proton interaction to others. Nuclear forces are symmetrical in neutrons & protons. i.e. the force between two protons are the same as those between two neutrons. This identity refers to the magnitude as well as the spin dependence of the forces.

Now we can see the following reaction



The capture of photons can effect the production of mesons by an electromagnetic interaction, decay electromagnetically since these processes involve no change of strangeness.



It is found that only two assumptions are in agreement with theoretical & experimental facts, notably the equality of the forces between two like & two unlike nuclear particles in the singlet state. These assumption are either(1) that nuclear

particles interact only with neutral mesons(neutral theory) or(2) that they interact equally strongly with neutral, positive & negative mesons(symmetrical theory). It is obvious that the part of the force which does not depend on the spin of the nuclear does not fulfill any useful function in the theory. The force between proton & neutron are result from the transfer of positive meson from the former to the latter or a negative meson in the opposite direction. So there vector sum of component of isotopic spin of these particle must be zero.

The charges on charged mesons must be equal in their magnitude.

Charge independence of nuclear forces demand the existence of π^0 meson as amongst the same type of nucleons (p-p) or (N-N). This force demand the same spin & orbital angular momentum. Positive pions are not able to surmount the nuclear coulomb barrier & there fore undergo spontaneous decay while negative pions are captured by nuclei. The exchange of a pion is thus equivalent to charge exchange. we can think of nucleons as exchanging their space & spin co-ordinates In the neutral theory, therefore neutron & protons are completely equivalent & indistinguishable as far as the associated meson fields are concerned. Such particle decay into two gamma rays. These gamma rays are π^0 - rest systems are emitted in opposite direction & therefore spin π^0 must be Zero as the spin of photon is unity. It is evident from the nature of the products that neutral mesons decay by the electromagnetic interaction while charged pions decay by strong & weak interaction both. It means that neutral mesons constituents responsible for the electromagnetic interaction. We know that neutron & proton can change into one another by meson capture. Protons & neutron can transform into each other by capture of positive & negative pion respectively, or get transform into the same particle through neutral meson interaction. During these

transformation either an emission or an absorption of meson is essential. The attraction between any nucleons can arise from the transfer of a neutral meson from one nucleon to the other. If the meson were assumed to be charged (positive or negative) the resulting force between nuclear particles turned out to be of the exchange type which had been successful in the interpretation in nuclear physics. The mesons must obey Bose statistics because they are emitted in the transformation of a neutron into a proton (or vice versa) both obey Fermi statistics.

Important point--

1. The deuteron does possess measurable properties which might serve as a guide in the search for the correct nuclear interaction. The mass number of deuteron A is very minimum. From the findings we must regard the deuteron as loosely bound. The deuteron consists of two particles roughly equal mass M , so that the reduced mass of the system is $1/2M$. The deuteron has spin $T=1$, the neutron & proton spins might be a parallel combination. The magnetic moment of deuteron will, therefore, be the sum of magnetic moments of proton & neutron. According to conclusion, as the range becomes larger in deuteron nucleus & they become more unstable.

*When light nuclei of hydrogen atom come within the range of nuclear force they can fuse together to form helium nucleus. In this process (fusion) the range is not effected. However the force is twice of the hydrogen's nuclear force & so on. Further the energy required to bring nucleons inside the range of force is twice of the rest energy of hydrogen nucleus. In other words, we can say that the minimum energy required to form a helium nucleus is twice of the hydrogen rest mass energy. The mass of He atom (alpha-particle) is equal to the four times of the hydrogen atom, so that the nuclear force of helium nucleus is four times stronger than that of hydrogen atom. It means that the He atom is more

stable than hydrogen atom. Because the binding energy of helium atom is larger than that of hydrogen atom, so in this process large energy is released. (Here the negative pion has a significant role to produce nuclear interaction. While charge pions maintain the character of proton.) Now, as the hydrogen nucleus converted into helium nucleus, these are happened when resonances of nucleons is in excited state. The scattering cross-section can be interpreted by assuming that in the strong interactions the total isospin is converted as well as the third component. The total isotopic spin of the system is $3/2$. The ratio of radius & range is about 2:1. It shows that those nuclei has maximum number of nucleons are most stable than less mass number nuclei.

❖

2*. Also, strongly, the velocity of light depends upon the range of the nuclear force. The velocity of light is equal like photons, lack mass & force carrying particles of other forces like strong force. Because the range is variable, then the velocity of light must be variable. As, velocity of light = Range of nuclear force (distance travel by meson) / Life time of resonances. In this relation we can see that the velocity of light must be variable. It is clear that the fundamental particles are not wholly independent. The neutron is observed to change spontaneously into a proton. Neutron decay takes on the average some thousands of seconds for free neutron, whereas within a nucleus the characteristic time between nucleon-nucleon collisions is 10^{-24} seconds. For a satisfactory

picture it is often enough to think of the nucleus as a grouping of protons & neutrons interaction, with the appearance or disappearance of photons. One should be noted that this relation hold only inside the nucleus. Out side the nucleus the evident is to be contrary. It is a fact no body (even mesons or gamma rays) can have velocity greater than the velocity of light. From this formula we can find the nuclear force acts between the pair of nucleons & does not influenced by the presence of neighboring nucleons. It is necessary that any one particle must brings the velocity of light. We know that the nuclear force is short ranged. Out side of the range it is repulsive.



*Range of nuclear force :- To show that the range of force is related to the mass of exchanged particle, assume that the π^0 -meson is contained virtually in a proton. If this virtual particle travels with the velocity light as might be expected for a field particle, then greatest distance the meson could travel in this time also known as range of the pion exchange force.

3. It would seem that in a nucleus consisting of the many nucleons the binding energy per nucleon should increase with the increase of the mass number A . In reality evidence is to contrary, the binding energy per nucleons decreases with increasing mass number A , The binding energies of the different nucleon placed at various depths are not identical but depend upon the states of their actual binding in the potential well. The binding energies of the different nucleon placed at various

depths are not identical but depends upon the state of their actual binding in the potential well. The range also depends upon mass number A & binding energy. we know that the atomic mass number A is approximately equal to a twice the atomic number Z . For the light & intermediate nuclei. It shows that light nuclei prefer to add nucleons is n-p pair. i.e there is a strong interaction between neutrons & protons. The range of nuclear force depends on the mass number A & the velocity of light depends on the range, so it is obviously thought that the spin & velocity of light depends on the mass number A of the nucleus & spin is zero or an integer for A even & is an odd half integral for A odd. The total rest mass energy also depends on the mass number A . For increasing of rest energy, we must increase the mass number A . obviously, the rest mass energy must be depends on the radial distance.

. This is purely a quantum mechanical effect. If the mass number A increases the range decreases, & the force are stronger. This binding energy displays saturation effect. This property of the nuclear force can be explain in term of exchange nature of nuclear force. It should be noted that nucleons attract each other strongly only if they are in same orbital state. This formula prove the pauli hypothesis. This formula usually attributed to the effect of higher-order interactions in which two or more mesons are simultaneously transmitted between the nucleons.

4. The velocity of light depends on the wavelength of it constituents, If the particles has longer wavelength then the range decreases & therefore force is stronger. we can find the effective range of nuclear force in terms of the Compton wave length of pi—meson. We know that different (variable) constituents (color particles) has different wavelength, so it is obviously thought the velocity light must be variable. Evidently, if meson interacted with nucleons strongly enough to be

responsible for the nuclear forces. Its mean free path within the nucleus should be about the same as is that of a nucleons. Then one question must be arise, How & why the velocity of light vary in free path or in vacuum. So far this problem is concern, the velocity of light influenced by its internal matter, which must have different value. However it may be possible that the velocity of light equal at all ranges within the nucleus. . The forces responsible for binding the individual particle inside the nucleus must therefore be exceptionally strong.If the particles has motion then the material body has physical significance otherwise not. It means the force between elementary particles depends on the velocity of the body as well as mass of body. It should be remarked that the particles travels with velocity of light are not a conservable quantities. In quantum theory, every field must be quantized. These quanta produce a field, which is responsible for different forces.

5.The emission of a charged meson will be accompanied by a change of charge of the emitting nuclear particle, Thus a neutron can only emits a negative meson or absorb a positive meson & will thereby be transformed into proton. When we consider the emission of one meson by a nuclear particle & reabsorbtion by another.It is obvious that in this way no force will be obtain between two nuclear particle of same kind i.e. two neutrons or two protons. For the same kind particles, there is neutral meson is responsible for interaction.This solution would make the interaction caused by neutral meson alone, Since for unlike particles the charged mesons given an additional contribution, while for like particles they do not, the total interaction will not be the same³ for like & unlike particles in S-first state.This will lead to forces between a neutron & a proton.The negative & positive charge

meson comes to close together, the can neutralize each other then the force between neutron & proton come into play. So obviously we can say that only neutral meson plays important role in charge independent nuclear force. The mesons(positive & negative) can be absorbed by the nucleus of an element or it may be combine with the another meson then the sum of the masses of these mesons converted into energy. This process is called annihilation of matter. Before this process one positive meson & one negative meson unite to make neutral particles called K^0 meson. The process of construction & destruction has proved very help full is considering the origin of universe. The neutral K^0 mesons is a stable particle but stability last for a small time. Its half life is of the order of micro seconds. This particle is an essential constituent of the nucleons of all elements.We know that neutral—meson decay into two photons & never into three photons. It is clear that the neutral pions has been produced by bombarding hydrogen & deuteron with high energy photons. Gamma rays has sufficient energy to maintain energy of nucleons then the nucleons produce the neutral mesons.One can speak of the meson field associated with a proton(or a netron) because the nature (charge) of the nuclear particle does not change by emitting or absorbing a neutral meson.It has developed a theory of nuclear forces in which neutral way the equality of the forces between like & unlike nuclear particles.The theory involving charged meson only giving no forces between two like nuclear particles. It is obvious that the negative meson & positive meson gave the symmetrical force between protons & neutrons & the interact equally strongly with these meson. An alternative way of explaining this equality is to assume interaction with neutral meson only. Then the charge of the nuclear particles(whether it is a proton or a neutron) becomes entirely irrelevant & the equality of forces follows immediately. This alternative is discussed in present paper.

6. According to the Pauli principle only two neutrons & two protons will be found in the same orbital state. Therefore it is possible to find four nucleons strongly bound or Alpha particle structure, also confirmed by binding energy curve. The extraordinary stability of the alpha shows that the most stable nuclei are those in which the number of nucleons & protons are equal. We can find it from this formula. It is obviously thought that the full charge independence for any system in which the number of neutrons equal to the number of protons. From conclusion, we get, Number of photons = number of nucleons = 2(number of neutral mesons). The discovery of the neutral meson & the fact that charge independence is now consistent with all nuclear data, confirm fully the use of the symmetric meson theory, containing positive, negative & neutral mesons described by three wave functions. With the form of Yukawa potential for scalar mesons, it is easy to see that the pi-meson can not be scalar. This theory proves this argument.

Change of Law--

Since there is no requirement for the conservation of Pions so there is no conservation law in rest mass energy & even in the universe. This formula shows that there is no meaning of the word 'constant'. There is no conservation law controlling the total number of Kaons or mesons. The energy of formation of mesons comes from binding potential (which has the energy to formation of meson for a long time), but when this potential has not enough energy, the production of pions ends & nuclear force does not exist.

*The life time of any radioactive substances depends on the total number of pions production and other particles production. Pions are commonly formed in the decay of kaons, hyperons & resonant states. It should be noted that pions are formed only at high energy. Because of their short life time

of neutral mesons move only a few atomic diameters before they decay (so that it influences few neighboring nucleons) & thus are not affected by the matter through which they pass & thus nuclear force works properly. It should be also noted that in the whole universe there is only mass will be conserved and energy will be destroyed, then the mass will not change into energy.

*It is enough to think that π^- mesons which form a nuclear cloud around the individual

nucleons & are in a virtual state get their requisite rest mass energy from the incident particle & are released from the nuclear binding potential. Nuclear binding potential compensates the rest mass energy. It produces enough energy to maintain the rest mass energy for production of mesons. Since the rest mass energy of π^- mesons is about 139 MeV, the threshold energy for gamma rays to produce the rest mass energy of these particles should be high. But, if protons projectiles are used to produce mesons, it requires a large threshold as a particle with mass retains some energy in the collision. It should be remarked that the binding potential is independent of spin & range of the particle, when they compensate the rest energy. The energy required to pull out the nucleons from the nucleus is less than half of rest mass energy. The slow motion neutron plays this role. Similarly if the nucleus brings (from binding potential) sufficient energy for the existing of nuclear force. It maintains stability. In order to approach particle to within short range or closer the energy of the approaching particle should be very high.

❖ Since there is no limitation of formation of Mesons even in strong interaction. This is due to high energy photons (γ - rays) then this cyclic chain should be possible

$\pi^0 \rightarrow \gamma$ rays

$\gamma + d \rightarrow d + \pi^0$

This reaction shows that the kinetic energy as well as potential energy of nucleon in the nucleus will be over and above of the rest mass energy. In these phenomena the total charge of fundamental particles are conserved.

- ❖ It is reasonable to assume that the nuclear force between two protons has the same characteristic as that between neutron & proton. The argument about short range forces involves both proton-proton & neutron - proton forces. The main difference between proton & neutron seems to be the electric charge, & the nuclear force apparently does not arise from charge. We assume therefore that the potential between two protons is confined within some short range as before, although the value of range need not necessarily be the same.

Finally-- some of the peculiarities of nuclear forces can enumerated as follows—

- (a) short range character
- (b) Large strength, the nuclear potential energies are quite large.
- (c) Exchange character & saturable nature.
- (d) Dependence on spin
- (e) charge independence & charge symmetry.

References—1. Elementary nuclear theory. H.A. Bathe

2. Nuclear physics, Roy & Nigam

3. Nuclear physics, Srivastva

Impact Fatigue Behaviour of fully dense Alumina Ceramics with Different Grain Sizes

Manoj Kumar Barai, Jagabandhu Shit, Abhijit Chanda, Manoj Kr Mitra

Abstract: Impact fatigue behavior of fully dense alumina as we studied in this work. The effect of grain size on the impact fatigue characteristics has been found out using a simple impact fatigue test set-up. Low grained alumina was prepared using optimized slip casting technique where as higher grained samples were made following conventional powdered metallurgy technique (Compaction through isostatic pressing and subsequent solid state sintering. All though the mechanical behavior (e.g.hardness, toughness) was better in fine grained alumina, It was more susceptible to impact fatigue. Some of the sampled in higher grained alumina samples were little bit elongated in shape with aspect ratio close to 2-2.5. Fractography revealed that crack propagation was predominantly mixed mode. The elongated grains promoted bridging across crack front and caused higher resistance to fatigue.

Index Terms— dynamic, element, factor, finite, impact, intensity, point, quarter, stress

1 INTRODUCTION

OWING to high hardness, compressive and extremely good corrosive resistances alumina is one of the mostly used turbo materials. In last few decades few few of studies have been done on the dependency of grain size on impact fatigue behavior of alumina. B. K. Sarkar and T.G.J Glinn (1969) studied the impact fatigue of an alumina ceramic and exhibit fatigue behaviors, having a high stress plateau followed by progressively increasing endurance with decrease in applied impact energy. S Maity and B.K.Sarkar (1994) studied the impact fatigue of a porcelain ceramic and showed a definite fatigue behavior with increasing endurance in decreasing impact energy levels and cumulative residual stress is suggested to explain the fatigue behaviors. S. Maity, D. Basu and B.K. Sarkar (1994) studied the fatigue behaviors of fine-grained alumina under repeated impact loading and found that the fatigue resistance parameter is 17.12 while the endurance limit is around 270Mpa which is about 38% of the single impact strength of the material and also found that fatigue cracks are trans-granular near the crack initiation region, the rest being inter-granular. Manabu et al (2002) studied the material response to particle impact during abrasive jet machining of alumina. A relatively smooth face can be produced when silicon carbide (GC) abrasive is employed.

The fatigue behaviour of fine-grained alumina hip joint heads under normal walking load has also been reported by Basu et al (2005) and found that the alumina femoral

heads have successfully withstood 10^7 cycles at maximum walking stress of 17.2 KN, which is equivalent to a body weight of 400Kg. The femoral heads didn't exhibit any sub-critical crack growth at the maximum walking load of 10KN, indicating the quasi-infinite performance life in-patient up to body weight of 250 Kg.

In recent past as new grade of alumina powder is available with very small particle size and new processing routes are involved for new generation alumina products, which are coming in big way. Most of these new grades of alumina are of smaller grain size. It has been found from the literatures that not too many studies on impact fatigue behavior of alumina with fine grain size (particularly sub-micron grain) have been done so far. So in the present work we have developed a machine, and using that we have seen the effect of grain size on impact fatigue behavior of alumina ceramics.

2 OBJECTIVES

The objectives of the present study are as follows.

- i) To develop an Impact Testing Machine for carry out the Experiment
- ii) To see the Effect of Grain Size on Impact Fatigue behavior
- iii) To study the dynamic stress intensity factor of Alumina under impact loading,

3 EXPERIMENTAL METHODS

Sample preparation : Two grades of high purity commercially available alumina powders with (i) average particle

- Manoj Kumar Barai is currently pursuing an integrated Ph.D. program in Jadavpur University, Kolkata-32, West Bengal, India. E-mail: manojmithu@yahoo.co.in
- JagabandhuShit is currently pursuing Ph.D program in Jadavpur University, Kolkata-32, West Bengal, India. E-mail: Jagabandhushit@rediffmail.com

size of 180 nm, purity 99.99 % and (ii) average particle size of > 0.4 μm, purity 99.8 % have been used for this study. The powders were used to obtain two different grained dense compacts with average grain size of 0.4μm (Fig 1) and 4 μm (Fig. 2). Alumina of different grain sizes was prepared from different powders using different methods as discussed below.

- a. Samples were prepared in two different routes :
 (a) Iso-pressing followed by sintering and (b) slip casting followed by pre-sintering and sintering
 The process flow-sheets are given below.

The density of the sintered samples was measured using water-immersion method. The sintered and ground samples were then polished for measuring various mechanical properties like hardness, toughness etc. With the conventionally prepared (iso-pressed and sintered) samples, hardness was found to be 16.0 GPa (average) and toughness was 3.5-4 Mpam^{1/2} . The density was found to be around 99% of the theoretical value.

In case of alumina made with slip casting process, density was in excess of 98%, hardness was as high as 28 GPa and toughness was comparable with that of the other grain size.

Microstructure :

SEM photographs of the alumina samples of different grain sizes are shown below (Fig 3&4).

Fig 3 Alumina samples sintered at 1275C following slip casting with a grain size of 0.4 μm

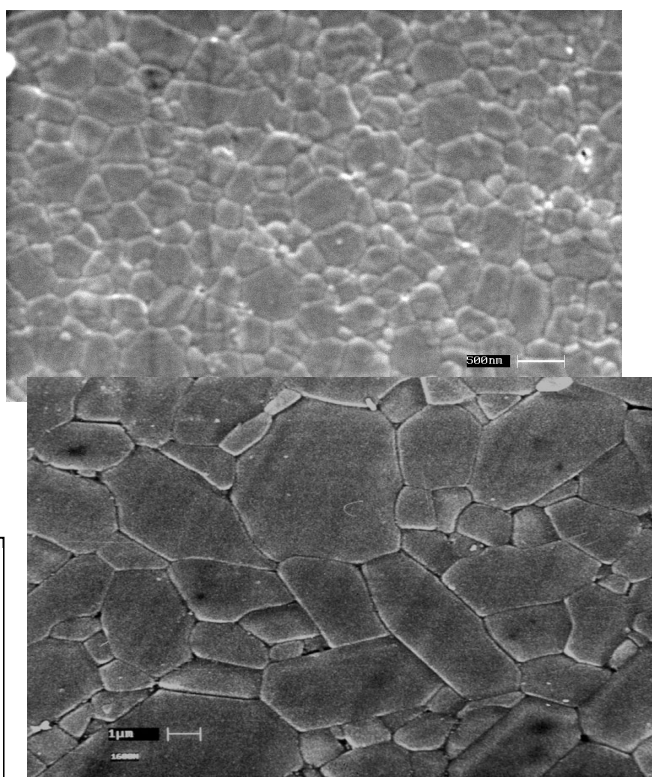


Fig 4 Alumina samples sintered at 1600C following iso-pressing and pre-sintering with grain size 4 micron

With alumina with smaller grain size the grains are mostly equiaxed where as for alumina with average grain size, grains are little bit elongated in shape which enhances grain bridging during crack propagation. A small set up for measuring the impact fatigue behaviour has been developed with a swinging pendulum of length of 54 cm and a weight in the form a spherical ball of diameters 36mm. The impact load is given to the test specimen perpendicular to its axis, which is rigidly fixed between two

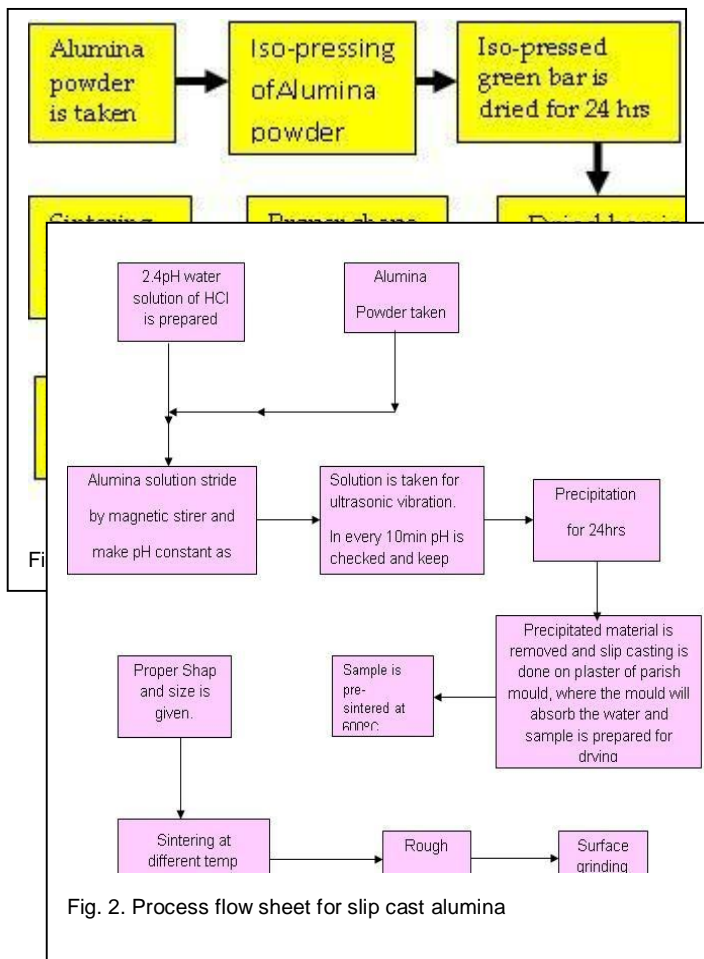


Fig. 2. Process flow sheet for slip cast alumina

supports in a configuration resembling that of a Charpy test specimen. The bob (concentrated mass) is fixed at the end of the bar (swing arm), which is supported at almost frictionless hinge. The hinge is fixed inside the bearing,

which is mounted in the bracket plate. The angular movement is given to the set (drum & bob) by the cam mounted on the low speed motor shaft. On the tensile surfaces of the beam specimens made of two different grained alumina, notches were created with different dimensions as mentioned below.



Fig 5 Experimental setup for impact fatigue testing

Dynamic stress intensity factors were found out using straight cracks on the tensile surfaces. Depths of the straight cracks were varied. Depth of notch is given by "a" and width of the specimen is given by "w". "a/w" is the non dimensional crack depth. This was done for both the batch of alumina samples. It was not possible to produce exactly similar notches but it was ensured that the depths were comparable.

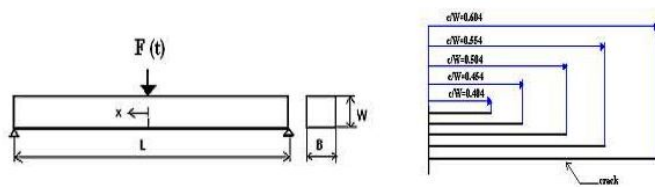


Fig 6 (a) Three point bend specimen with straight crack, (b) The range of crack depths in three point bend configuration.

4. RESULTS & DISCUSSION

To perform the test we have taken samples of different grain size such as 0.45µm and 4µm and it was observed that the fatigue of ceramic samples of smaller grain size was poorer than the higher grain size alumina. From the

experimental results it was observed that with the increase of crack length the no of strike required for breaking the sample was decreasing for both the batch of samples. But it was observed that even if the crack dimension was more the no of strike required for breaking the sample of grain size 4µm was more than the sample of grain size of 0.45µm. This was probably due to the shape of the grains in higher grained samples which enhanced the scope of grain bridging. In case of number of impacts required for breaking, there was variation in the pattern as well.

Table 1: Number of strikes reqd to break vs. a/w for alumina with grain size of 0.45 micron

EFFECTIVE LENGTH OF THE SAMPLE: 50mm TESTING WITH SMALL BALL & SMALL CAM a=Notch depth, w=Sample width					
GRAIN SIZE	NOTCH DEPTH	NO OF STRIKE	C/S OF THE SAMPLE	a/w	AVG. STRIKE
0.45µm	0.228mm	1	5x5 sq. mm	0.0456	1
0.45µm	0.228mm	1	5x5 sq. mm.	0.0456	
0.45µm	0.228mm	1	5x5sq. mm.	0.0456	
0.45µm	0.125mm	2	5x5sq. mm.	0.025	2
0.45µm	0.125mm	2	5x5sq. mm.	0.025	
0.45µm	0.125mm	1	5x5sq. mm.	0.025	
0.45µm	0.05mm	17	5x5sq. mm.	0.01	65
0.45µm	0.05mm	59	5x5 sq. mm.	0.01	
0.45µm	0.05mm	119	5x5 sq. mm.	0.01	

Fig. 7. Variation of number of strike to break with non-dimensional crack length.

From Fig. 7 it is evident that with alumina having lower grain size, samples got broken with almost a single blow in case of higher a/w (>0.025) while for the lowest a/w (=0.01), there was a large scatter in the data. It shows that with decrease in non-dimensional crack length (a/w), influence of other intrinsic material features (e.g. presence of other micro-structural discontinuities) got pronounced resulting in high scatter from around 20 to 120. It was also observed that a second degree polynomial equation nicely fits the data with high (>95% fitting) percentage fitting.

Table 2. No of strikes to required to break vs a/w for alumina with grain size of 4 micron.

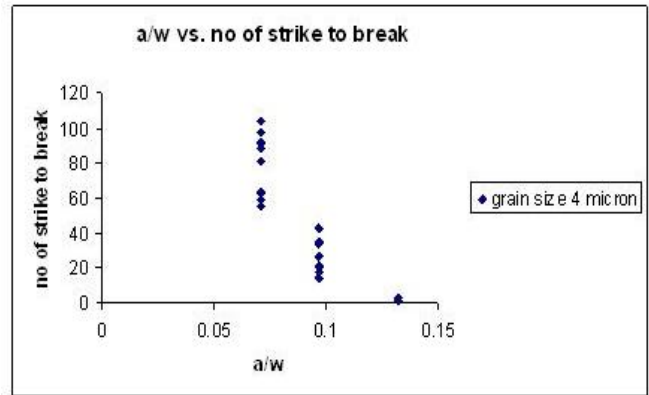
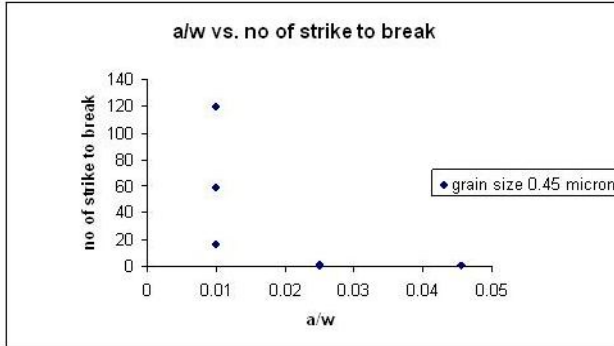


Fig. 8. Variation of number of strike to break with non-dimensional crack length

EFFECTIVE LENGTH OF THE SAMPLE: 50mm
TESTING WITH SMALL BALL & SMALL CAM
a=Notch depth, w=Sample width

GRAIN SIZE	NOTCH DEPTH	NO OF STRIKE	C/S OF THE SAMPLE	a/w	AVG. STRIKE
4µm	.662mm	2	5x5	0.1324	2
4µm	.662mm	2	5x5	0.1324	
4µm	.662mm	3	5x5	0.1324	
4µm	.662mm	2	5x5	0.1324	
4µm	.662mm	2	5x5	0.1324	
4µm	.662mm	2	5x5	0.1324	
4µm	.485mm	15	5x5	0.097	28
4µm	.485mm	18	5x5	0.097	
4µm	.485mm	27	5x5	0.097	
4µm	.485mm	34	5x5	0.097	
4µm	.485mm	35	5x5	0.097	
4µm	.485mm	21	5x5	0.097	
4µm	.485mm	43	5x5	0.097	80
4µm	0.356	55	5x5	0.0712	
4µm	0.356	63	5x5	0.0712	
4µm	0.356	81	5x5	0.0712	
4µm	0.356	59	5x5	0.0712	
4µm	0.356	92	5x5	0.0712	
4µm	0.356	104	5x5	0.0712	
4µm	0.356	88	5x5	0.0712	
4µm	0.356	97	5x5	0.0712	

For the samples with higher grain size, scatter was comparatively low, even the samples with a/w as high as 0.07 (even bigger than the largest a/w in the previous case) could withstand around 80 impacts (average) before fracture. With another set of specimens having a/w 0.097 (almost double the value of maximum a/w in the previous case), average number of impacts sustained before fracture was around 30. Only one set having high a/w (almost three times the maximum a/w of the previous case) got broken with single impact.

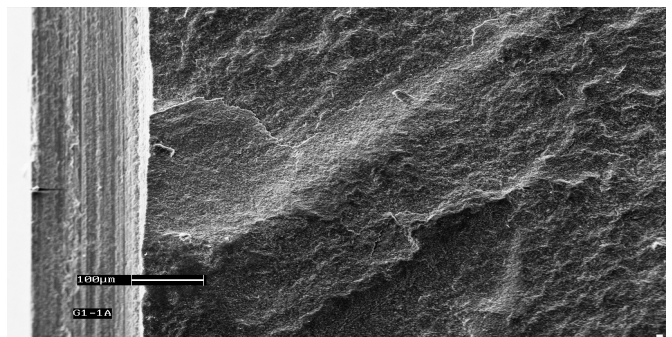


Fig 9 Cracks propagating (sideways) from the existing notch

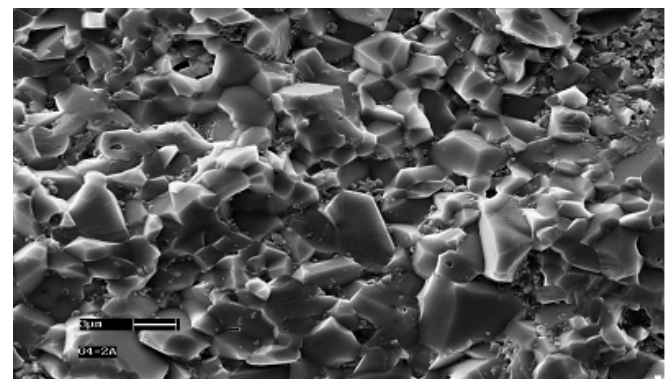


Fig. 10 Mixed mode of fracture in higher grained sample

From the SEM micrograph (Fig. 9), it is clear that the cracks started propagating from the sides of the notches and it followed a tortuous path. The crack apparently followed a new plane after reaching a particular point or obstacle as observed on the fractured surface. In few other cases, there was apparently a crack initiation site (just at the tip of the notch) surrounded by semi-circular arc-shaped zone. Fig. 10 shows that there was a mixed

mode of crack propagation: both trans-granular and inter-granular. From the data available so far, it is evident that higher grained alumina samples showed superior impact fatigue behaviour. The exact reason behind is not yet very clearly understood however a simple analysis of the two microstructures reveal that in case of higher grain size the shape was a little bit elongated in contrast with nearly equiaxed shape observed with sub-micron grained alumina. This caused grain bridging retarding partially the propagation of cracks along grain boundary in case of larger grained alumina. Furthermore with higher grain-size chance of crack-branching could be more that could reduce the energy available for the propagation of the main crack front.

5. CONCLUSIONS:

From the study it is evident that alumina with higher grain size showed superior impact fatigue behaviour in comparison with that of sub-micron grained alumina. Even with higher non-dimensional crack length (a/w), alumina specimens with bigger average grain-size withstood higher number of impacts prior to fracture. The superior impact fatigue behaviour was due to higher resistance to fatigue crack growth owing to grain bridging.

ACKNOWLEDGMENTS

We acknowledge sincere help to the staff members of Mechanical Engg Dept. & SBSE, Jadavpur University, Kolkata. One of the authors (M.K. Barai) sincerely acknowledge constant support from the HOD, Director and Principal of Future Institute of Engg and Management.

REFERENCES

- [1]. Anderson, T.L., Fracture Mechanics: Fundamentals and Application. 1995, CRC Boca Raton
- [2]. Barsoum R.S. Triangular quarter- point elements as elastic and perfectly plastic crack tip elements. In International Journal for Numerical methods in Engineering. (11),pp. 85-98 (1977)
- [3]. Bohme W, Kalthoff JF. The behavior of notched bend specimens in impact testing. International Journal of Fracture 1982;20(4): R139-43
- [4]. Chen Y.M. Numerical computation of dynamic stress intensity factors by a Lagrangian finite-difference method. Engineering Fracture Mechanics 7(4), 653-660 (1975)
- [5]. Enderlein, M., Ricoeur, A., Kuna, M. Comparison of finite element technique for 2D and 3D crack analysis under impact loading. International Journal of Solids and Structures 40(13-14),3425-3437, 2004
- [6]. Isida M., Effect of width and length on stress intensity factors of internally cracked plates under various boundary conditions. International Journal of fracture, 7, 301-316 (1971)
- [7]. John R., Stress intensity factor and compliance solutions for eccentrically loaded single crack geometry. Engineering Fracture Mechanics (58) ½ pp. 87-96, 1997
- [8]. John R. and Rigling B., Effect of height to width ratio on K and CMOD solutions for a single edge cracked geometry with clamped ends. Engineering Fracture Mechanics (60) No. pp 147-156, 1998
- [9]. Kishimoto K., Aoki S. Sakata M., Dynamic stress intensity factors using J-integral and finite element method. Engineering Fracture Mechanics 13(2),387-394, 1980
- [10]. Maity S. Sarkar B.K., "Impact fatigue of porcelain ceramic" International Journal of Fatigue" 1995; 17(2), 107-109.
- [11]. Meggiolaro M. A., Miranda. A. C. O., Castro J.T.P., Martha L. F., Stress intensity factors for branched crack growth. Engineering Fracture Mechanics. 7 2 (2005) pp. 2647- 2671.
- [12]. Nishioka T., Computational dynamic fracture mechanics, International journal of fracture 86 (1997) pp. 127-159
- [13]. Rokach I. V., On the numerical evaluation of the anvil force accurate dynamic stress intensity factor determination. Engineering Fracture Mechanics 70(2003) 2059-2074
- [14]. Rokach I. V., Estimation of the three-dimensional effects for the impact fracture specimen. Arch Mech Engg 1996 :43(2-3):241-252
- [15]. Rokach I.V., Mixed numerical-analytical approach for dynamic one point bend test modeling., International journal of fracture 130,L193-L200 2004
- [16]. Sinclair G. B., Messner T. W., Meda G., Stress intensity factors for deep cracks in bending. Engineering Fracture Mechanics Vol. 55 No.1 pp. 19-24, 1996
- [17]. Sinclair G. B., Meda G., Galik K., stress intensity factors for side-by-side edge cracks under bending. Engineering Fracture Mechanics Vol. 57 No.55 pp. 577-581,1997
- [18]. Weisbrod G., Rittel D., A method for dynamic fracture toughness determination using short beams. International Journal of Fracture (104) pp.89-103 2000

Distributed Generation Planning Optimization Using Multiobjective Evolutionary Algorithms

Mahmood Sheidaee, Mohsen Kalantar

Abstract— In this paper, a method to determine the size - location of Distributed Generations (DGs) in distribution systems based on multi objective performance index is provided considering load models. We will see that load models affect the location and the optimized size of Distributed Generations in distributed systems significantly. The simulation studies are also done based on a new multi objective evolutionary algorithm. The proposed method has a mechanism to keep the diversity to overcome the premature convergence and the other problems. A hierarchical clustering algorithm is used to provide a manageable and representative Pareto set for decision maker. In addition, fuzzy set theory is used to extract the best solution. Comparing this method with the other methods shows the superiority of proposed method. Furthermore, this method can easily satisfy other purposes with little development and extension.

Index Terms— Distributed generation, Distribution systems, Load models, Strength Pareto Evolutionary Algorithm.

1 INTRODUCTION

Optimization was used to reconstruct electricity industry and looked for the best location for distributed products. Development in technology and client requirements to have cheap electric power and reliable one caused more motivation in distributed generation. Discussion about reliability and maintaining prevent the penetration of DG resources in the distribution systems.

In [1] one approach was described based on genetic algorithm for multistage planning of distribution systems optimizations. In this work, it's expressed as a mathematical model and algorithmic one and also tested with real systems. In [2] – [5], it was studied on load models that are usable for power flow and dynamic studies. This study was done on load models depended on frequency or voltage. During the recent years, studies on evolutionary algorithm have shown that these methods don't have the difficulties of classical methods. In principle, multiple Pareto optimal solutions can be found in one single run.

This paper has discussed on load model effects in location and size planning and distributed generation optimization. We can see that the load models affect on location and size planning of DGs in distribution network. For the purpose of studying on load models, its delivered location and size planning for single DG, its assumed that the regarded DG has enough capacity. The proposed method is general and it can be used for case of multiple DG in the network with increasing some variables.

This paper also suggested a new Strength Pareto Evolutionary Algorithm (SPEA) based approach for solving the problem. The diversity-preserving mechanism embedded in the search algorithm makes it effective in exploring the problem space and capable of finding widely different solutions. A hierarchical clustering technique is implemented to provide a representative and manageable Pareto-optimal set. Also, a fuzzy-based mechanism has used the best solution for extraction.

2 LOAD MODELS AND IMPACT INDICES

To determine different load model effects on distributed generation planning, 37-bus distribution system will be studied (appendix 1)[7]. The effect of load models depends on voltage, means residential, industrial and commercial, will be studied in different planning scenarios. Load model defined as followed.

$$P_i = P_{0i} |V_i|^\alpha / Q_i = Q_{0i} |V_i|^\beta \quad (1)$$

Where P_i and Q_i are active and reactive power at bus i , P_{0i} and Q_{0i} are active and reactive power operating point in bus i , V_i is voltage in bus i and α and β are active and reactive power exponents. In a constant power model conventionally used in power flow studied $\alpha = \beta = 0$ is assumed. The values of the real and reactive exponents used in the present paper for industrial, residential and commercial loads are given in Table 1 [7].

TABLE 1
EXPONENT VALUES

Load Type	α	β
Constant	0	0
Industrial	0.18	6.00
Residential	0.92	4.04
Commercial	1.51	3.40

During studying residential, it's assumed that 38-bus systems just has residential load. It's assumed that for industrial and commercial load, all loads are a kind of industrial and commercial. In real situations, loads aren't exactly residential, commercial and industrial, so the mixture load class should be foreseen for distribution system. There are different ideas for studying DG effects in distribution systems. One of this idea is different

indices evaluation on the purpose of effect description on distribution system because of DG during maximum power production. These indices are

1) *Active and Reactive Power Loss Indices (ILP and ILQ)*:

$$ILP = \frac{P_{LDG}}{P_L} \times 100 / ILQ = \frac{Q_{LDG}}{Q_L} \times 100 \quad (2)$$

Where P_{LDG} and Q_{LDG} are total loss of active and reactive power distribution system with DG, P_L and Q_L are total loss of active and reactive power of total system without DG in the distribution network.

2) *Voltage Profile Index (IVD)*: One of the advantage of proper location and size of the DG is the improvement in voltage profile.

$$IVD = \max_{i=2}^n \left(\frac{|V_1| - |V_i|}{|V_1|} \right) \times 100 \quad (3)$$

3) *MVA Capacity Index (IC)*: This informational index gives information in the field of system necessities for promoting transmission line.

$$IC = \max_{i=2}^n \left(\frac{|S_{ij}|}{|CS_{ij}|} \right) \quad (4)$$

3 PROPOSED APPROACH

Recently evolutionary algorithm showed that this algorithm can be effective for removing old method problems [8]. The main element method of SPEA is

1) *External set*: It's a set of Pareto optimal solutions. These solutions were recorded externally and continuously be updated. Finally recorded solutions show Pareto optimal front.

2) *Strength of a Pareto optimal solution*: It is an assigned real value $S \in [0,1]$ for each individual in the external set. The strength of an individual is proportional to the number of individuals covered by it.

3) *Fitness of population individuals*: Fitness of each individual in population is the sum of the strengths of all external Pareto optimal solutions by which it is covered. The strength of a Pareto optimal solution is at the same time its fitness.

Algorithm is in the following steps [8].

Step 1) primary amounts: produce population and make empty external Pareto optimal set.

Step 2) updating external set: External Pareto optimal set is updated as following:

a) Search population for the nondominated individuals and copy them in the external pareto set.

b) Search external Pareto set for the nondominated individuals and emit all dominated individuals from the set.

c) If the amount of the individuals externally stored in the Pareto set exceeds a prespecified maximum size, reduce the set by means of clustering.

Step 3) Fitness assignment: Calculate the amount of fitness values of individuals in both external Pareto set and the population as follows.

a) Assign appropriate each individual's strength amount in external set. The strength amount is proportional to the number of individuals covered by that individual.

b) The fitness of each individual in population is equal to the sum of the strengths of all external Pareto solutions which dominate that individual.

Step 4) Selection: combine the population and external set individuals. Choose two individuals randomly and compare their fitness. Choose the best one and copy in a mating pool.

Step 5) Crossover and Mutation: do the crossover and mutation according to new population production probabilities.

Step 6) Ending: check the ending criteria if all things are being done finish the work else substitute the old population with the new one and go to step2.

In this paper, time searching will be stopped if the generation counter exceeds its maximum number.

In some cases, the Pareto optimal set is extremely big or has extra solutions. An average linkage based hierarchical clustering algorithm is used to reduce the Pareto set. We want to change P given set which its size exceeds the maximum allowable size N to P* set with size of N. Algorithm is such as following [8].

Step 1) Give primary amount to set C. each member of P means a distinct cluster.

Step 2) if the number of clusters $\leq N$, go to Step 5, else go to Step 3.

Step 3) Calculate all the pairs of clusters distance. The distance d_c of two clusters $C_1, C_2 \in C$ is given as the average distance between pairs of individuals across the two clusters

$$d_c = \frac{1}{n_1 n_2} \sum_{i_1 \in C_1, i_2 \in C_2} d(i_1, i_2) \quad (5)$$

Where n_1 and n_2 are clusters individuals of C_1 and C_2 . Function d shows Euclidian distance between i_1 and i_2 .

Step 4) Determine two clusters that have minimum d_c distance. Combine these clusters into a larger one. Go to Step 2.

Step 5) find centroid for each cluster and choose the nearest individual to the centroid as agent and emit other individuals from the cluster.

Step 6) Compute the reduced nondominated set P^* by uniting the representatives of the clusters.

As soon as having the Pareto optimal set of nondominated solution, the proposed approach presents one solution as the best compromise solution. Each objective function of the i -th solution is represented by a membership function μ_i defined as

$$\mu_i = \begin{cases} 1 & F_i \leq F_i^{\min} \\ \frac{F_i^{\max} - F_i}{F_i^{\max} - F_i^{\min}} & F_i^{\min} \leq F_i \leq F_i^{\max} \\ 0 & F_i \geq F_i^{\max} \end{cases} \quad (6)$$

For each nondominated solution, the normalized membership function μ^k is

$$\mu^k = \frac{\sum_{i=1}^{N_{obj}} \mu_i^k}{\sum_{k=1}^M \sum_{i=1}^{N_{obj}} \mu_i^k} \quad (7)$$

where M is the number of nondominated solutions. The best solution is the one that has more μ^k .

4 IMPLEMENTATION OF THE PROPOSED APPROACH

Because of Binary representation problems when search space has wide dimension, the proposed approach has been implemented using Real Coded Genetic Algorithm (RCGA). Decision variable x_i has real amount within limit of a_i and b_i ($x_i \in [a_i, b_i]$). The RCGA mutation and crossover operators RCGA is like this.

Crossover: A blend crossover operator (BLX- α) has been employed in this paper. This operator will choose one number randomly from the interval $[x_i - \alpha(y_i - x_i), y_i + \alpha(y_i - x_i)]$, where x_i and y_i are the i th parameter values of the parent solutions and $x_i < y_i$. Because of ensure the balance between exploitation and exploration from search space, $\alpha = 0.5$ is chosen.

Mutation: Nonuniform mutation was used here. In this operator, new amount x'_i of parameter x_i produced after mutation in t time.

$$x'_i = \begin{cases} x_i + \Delta(t, b_i - x_i) & \text{if } \tau = 0 \\ x_i - \Delta(t, b_i - x_i) & \text{if } \tau = 1 \end{cases} \quad (8)$$

$$\Delta(t, y) = y \left(1 - r \left(1 - \frac{t}{g_{\max}} \right)^\beta \right) \quad (9)$$

Where τ is a binary random number, r is a random number $r \in [0, 1]$, g_{\max} is maximum number of generations and β is a positive constant that is desirable. $\beta = 5$ is selected. This operator gives a value $x'_i \in [a_i, b_i]$ such that the probability of returning a value close to x_i increases as the algorithm advances. This makes uniform search in the initial stages where t is small and for later stages is so local.

5 MULTIOBJECTIVE BASED FORMULATION

Multiojective index for evaluating distribution systems operation on purpose of DG location and size planning with load models, considers all previously mentioned indices by strategically giving a weight. The multiojective index operation on basis of SPEA algorithm is according to (10).

$$IMO = (\sigma_1 \cdot ILP + \sigma_2 \cdot ILQ + \sigma_3 \cdot IC + \sigma_4 \cdot IVD) \quad (10)$$

These weights are because of giving the corresponding importance to each impact indices. Table 2 identifies used amount for the weights with regarding normal operation analysis [7].

TABLE 2
INDICES WEIGHTS

Indices	σ_p
ILP	0.40
ILQ	0.20
IC	0.25
IVD	0.15

Multiojective function (10) can be minimized with regarding to various operational constraints to satisfy the electrical requirements for distribution network. These limitations are:

1) **Power Conservation Limits:** The algebraic sum of all input and output powers, such as distribution network total losses and power generated from DG, which should be equal with zero. (NOL = no of lines)

$$P_{ss}(i, V) = \sum_{i=2}^n (P_D(i, V)) + \sum_{n=1}^{NOL} P_{\text{loss}}(V) - P_{DGi} \quad (11)$$

2) **Distribution Line Capacity Limits:** Transmission capability in each line should be equal with thermal capacity.

$$S_{(i,j)} \leq S_{(i,j)_{\max}} \quad (12)$$

3) **Voltage Drop Limits:** voltage drop should base on voltage regulation that DISCO gives.

$$|V_1 - V_j| \leq \Delta V_{\max} \quad (13)$$

If voltage and MVA limits in system buses for a particular size and location, except that pair for next generation, else this size and location will be ignored and rejected. Size and location should be had minimum IMO.

6 SIMULATION RESULTS

The multiobjective index based analysis is carried out on 37-bus test systems as given in the Appendix [7]. A DG size is considered in a practical range (0–0.63 p.u.). It's assumed that DG is operated at unity p.f.. This assumption has two reasons:

1) Usually when the DG has unity power factor, has maximum profit because the cost of active power is higher. Use at unity power factor cause to have maximum capacity.

2) Used models in this paper are simple and more attention is for voltage changes dependence of load models.

The using method hasn't been limited by DG models and it's general. First bus was chose as feeder of electric power from network and the rest buses are regarded as DG location. On all optimization runs, the population size and maximum number of generations were selected as 200 and 500, respectively. The Pareto optimal set maximum size includes 20 solutions. The crossover and mutation probabilities were selected as 0.9 and 0.01, respectively. For 37-bus system, variation of impact indices and IMO have been shown with DG size and location in figure 3-7 for constant, industrial, residential, commercial and mixed load models. The value of IVD for all load models is near zero. It shows that voltage profile improves with present DG.

We can see from Figs. (1) – (5) that the indices ILP, ILQ, IC and IMO achieve values greater than zero and smaller than one, indicating the positive impact of DG placement in the system. Fig. 1 shows that values of IC, ILP and ILQ for buses 2-4 as $IC < ILP < ILQ$ and for buses 6-8 like $ILQ < ILP < IC$. Figure 2 shows the value of optimum DG size, IMQ and its components for all buses for industrial load model. So load models affect on solutions.

Fig. 1. Impact indices and IMO with DG size-location pair for constant load

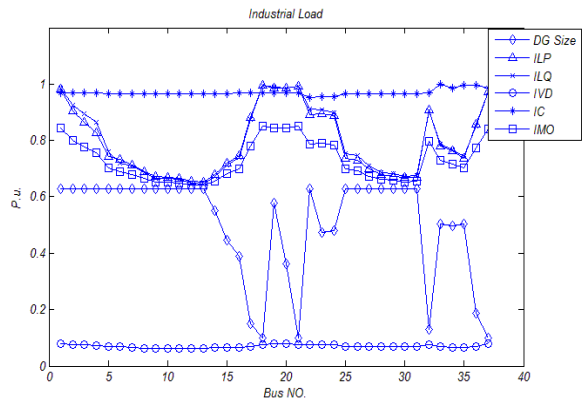


Fig. 2. Impact indices and IMO with DG size-location pair for industrial load

The solution obtained using constant power load models may not be feasible for industrial load. A similar and significant effect of load models can be easily be observed from the Figs. (3) – (5) for residential commercial and mixed load models. The differences in values of DG size, IMO and its components are significant, showing that the load models effects are important for suitable planning of size and location. Table 3 summarizes the optimal DG size-location pairs, IMO along with its components for each kind of load. From Table 3, the optimal size-location for constant load model (0.6299 p.u. – bus 14) is different with industrial load model (0.63 p.u. – bus 14) residential load model (0.4672 p.u. – bus 14) commercial load model (0.4419 p.u. – bus 14) and mixed load (0.5113 p.u. – bus 32). Similarly IMO and other effective indices for optimal DG location-size are different.

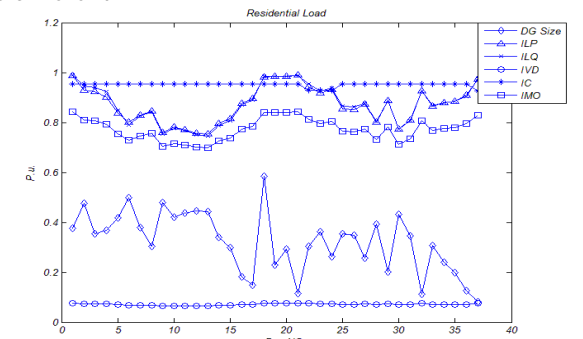
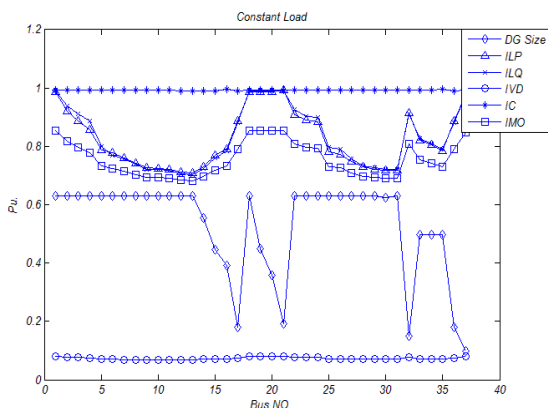


Fig. 3. Impact indices and IMO with DG size-location pair for residential load



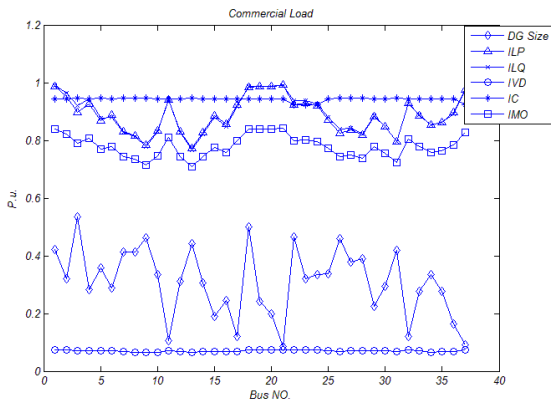


Fig. 4. Impact indices and IMO with DG size-location pair for commercial load

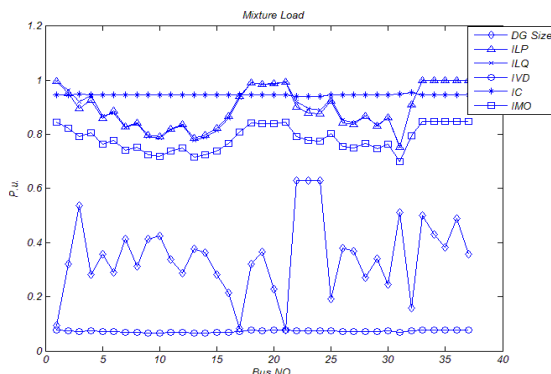


Fig. 5. Impact indices and IMO with DG size-location pair for mixture load

The probable DG location-sizes may be little (because of constraints), but the number of candidate solution are fairly large to suggest the application of SPEA. The differences in values of DG size, IMO and its components are significant for load models, showing that the load models effects are important for suitable planning of size and location. The values of Q_{LDG} and P_{LDG} related to optimal size –location for any kind of load model have been shown in table 4, although the values of Q_{LDG} and P_{LDG} for nonconstant load models (industrial – residential – commercial and mixture) aren't different but their difference is significant when compared to constant load model.

TABLE 3
IMPACT INDICES COMPARISON FOR PENETRATION OF DG WITH LOAD MODELS

Indices	Constant	Industrial	Residential	Commercial	Mixture
ILP	0.7078	0.6517	0.7459	0.7756	0.7526
ILQ	0.7035	0.6449	0.7383	0.7685	0.7551
IC	0.9913	0.9671	0.9570	0.9476	0.9478
IVD	0.0687	0.0634	0.0661	0.0653	0.0696
IMO	0.6823	0.6409	0.6952	0.7106	0.6994
Location	14	14	14	14	32
Size	0.6299	0.63	0.4672	0.4419	0.5113

TABLE 4
COMPARISON OF SYSTEM POWER LOSSES AT OPTIMAL LOCATION OF DG WITH LOAD MODELS

Load model	Optimal location	$P_{LDG} \times 0.01$ p.u.	$P_L \times 0.01$ p.u.	$Q_{LDG} \times 0.01$ p.u.	$Q_L \times 0.01$ p.u.
Constant	14	0.1499	0.2002	0.0991	0.1335
Industrial	14	0.1464	0.1671	0.0968	0.1112
Residential	14	0.1358	0.1604	0.0896	0.1066
Commercial	14	0.1166	0.1548	0.0767	0.1028
Mixture	32	0.1142	0.1588	0.0766	0.1056

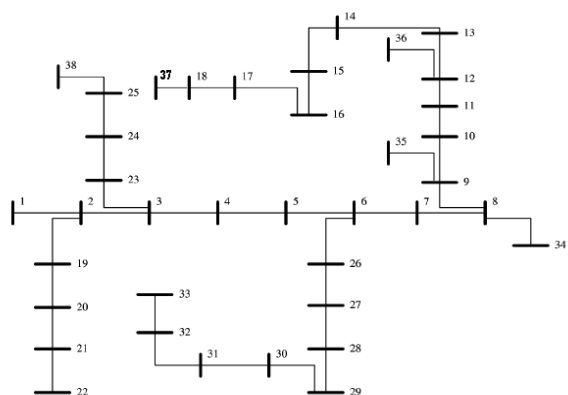
6.3 Conclusion

The general analysis includes load models is proposed for location-size of distributed generation planning in multiobjective optimization in distribution systems. The multiobjective criteria depends on system operation indices is used in this work. It was seen that while regarding load models, there will be changed in DG location and size. The overall value of multiobjective index (IMO) changed during charge model changing.

Also in this paper, we suggested a new method based on Pareto evolutionary algorithm and used for DGs location – size planning problem. This problem formulized as a multiobjective optimization problem, A diversity preserving mechanism for finding widely different Pareto optimal solutions was used. A hierarchical clustering technique is implemented to provide a representative and manageable Pareto optimal set without destroying the characteristics of the trade-off front and a fuzzy based mechanism is used for finding the best compromise solution. The result shows that the suggestive method for multiobjective optimization problem is useful, because multiple Pareto optimal solutions are found during simulation. Since the proposed approach does not impose any limitation on the number of objectives, its extension to include more objectives is a straightforward process.

APPENDIX

Fig. 6 shows the 37-bus test system.



REFERENCES

- [1]. V. Miranda, J. V. Ranito, and L. M. Proenca, "Genetic algorithms in optimal multistage distribution network planning," *IEEE Trans. Power Syst.*, vol. 9, no. 4, Nov. 1994, pp. 1927–1933.
- [2]. C. Concordia and S. Ihara, "Load representation in power systems stability studies," *IEEE Trans. Power App. Syst.*, vol. PAS-101, no. 4, Apr. 1982, pp. 969–977.
- [3]. IEEE Task Force on Load Representation for Dynamic Performance, "Bibliography on load models for power flow and dynamic performance simulation," *IEEE Trans. Power Syst.*, vol. 10, no. 1, Feb. 1995, pp. 523–538.
- [4]. IEEE Task Force on Load Representation for Dynamic Performance, "Load representation for dynamic performance analysis," *IEEE Trans. Power Syst.*, vol. 8, no. 2, May 1993, pp. 472–482.
- [5]. IEEE Task Force on Load Representation for Dynamic Performance,

Fig. 6. 37-bus test system

- mance, "Standard load models for power flow and dynamic performance simulation," *IEEE Trans. Power System*, vol. 10, no. 3, Aug. 1995, pp. 1302–1313.
- [6]. C. A. C. Coello, "A comprehensive survey of evolutionary based multiobjective optimization techniques," *Knowledge and Information Systems*, vol. 1, no. 3, 1999, pp. 269–308.
 - [7]. E. D. Singh, D. Singh, K. S. Verma, "Multiobjective Optimization for DG Planning With Load Models," *IEEE Trans. Power Syst.*, vol. 24, no. 1, Feb. 2009, pp. 427–436.
 - [8]. M. A. Abido, "Environmental/Economic Power Dispatch Using Multiobjective Evolutionary Algorithms," *IEEE Trans. Power Syst.*, vol. 18, no. 4, Nov. 2003, pp. 1529–1537.
 - [9]. C. M. Huang, H. T. Yang, and C. L. Huang, "Bi-Objective power dispatch using fuzzy satisfaction-maximizing decision approach," *IEEE Trans. Power Syst.*, vol. 12, Nov. 1997, pp. 1715–1721.
 - [10]. D. B. Das and C. Patvardhan, "New multi-objective stochastic search technique for economic load dispatch," *Proc. Inst. Elect. Eng.-Gen. Transm. Dist.*, vol. 145, no. 6, 1998, pp. 747–752.
 - [11]. M. E. Baran and I. M. El-Markabi, "A multiagent-based dispatching scheme for distributed generators for voltage support on distribution feeders," *IEEE Trans. Power Syst.*, vol. 22, no. 1, Feb. 2007, pp. 52–59.
 - [12]. E. G. Carrano, L. A. E. Soares, R. H. C. Takahashi, R. R. Saldana, and O. M. Neto, "Electric distribution network multiobjective design using a problem-specific genetic algorithm," *IEEE Trans. Power Del.*, vol. 21, no. 2, Apr. 2006, pp. 995–1005.

Marketing Of Asian Countries as Tourist Destination- Comparative Study of India and Malaysia

Dr Reshma Nasreen, Nguyen Toan Thang

Abstract - Tourism has gradually grown over the years as a full fledged industry. Many countries are gaining from this welcome change. The contributions of this sector to the country's coffers are sizable for some countries, while some countries have a long way to go. This research paper attempts to study the reasons of lack of optimal contribution of this sector in India and also forays into strategies that can be adopted to capitalize on the patterns prevalent in tourist behavior. A country like India with a commendable historical significance and size has not been able to garner as much of tourist attention because of certain factors. India has a lot of offerings to whip the appetite of an avid tourist, but the varieties have either not been promoted, or if promoted lack of associated services have not led to desired synergies. After identifying the gaps between the two countries (India & Malaysia), the paper puts forth the tourists' patterns of behavior through the data collected. The questionnaire has been administered to tourists in New Delhi and Agra (cities in India). Malaysia on the other hand has had a steady stream of tourists trickling down and benefiting its economy.

Key Words- ASEAN, Eco-tourism, Heritage Sites, MICE, Ministry Of Tourism (India & Malaysia). World Travel.

INTRODUCTION

Tourism plays an important role in the economies of a number of ASEAN as well as other Asian countries. Besides contributing to the national income, promotion of intra-regional tourism has other beneficial spillovers for trade and people-to-people contacts. The shared history and culture dating back to several centuries provides a base for tourism exchanges. A large number of tourists from Southeast Asia come to centre of Buddhist pilgrimage in India every year. With the rise of Indian middle class with higher purchasing power, India has also emerged as a big market for ASEAN countries as a source of tourists.

According to Amrik Singh, instructor at Department of Parks, Recreation and Tourism, University of Utah, "The Asia and Pacific region will be the focus of the worldwide tourism industry in the new millennium. Over the last decade, tourist arrivals and receipts rose faster than any other region in the world, almost twice the rates of industrialized countries." Statistics from

the World Tourism Organization (WTO) for 1996 show that tourist arrivals and receipts accounted for a 15.2% and 19.4% share of the world's total respectively, a significant increase from 1985 as shown in Tables I and 2 (WTO 1997a).

In his paper, Asia Pacific Tourism Industry: Current Trends and Future Outlook, Mr Amrik Singh points out, "that the rapid growth of the tourism industry has been attributed to a number of factors including among others, strong economic growth, increase in disposable income and leisure time, easing of travel restrictions, successful tourist promotion, and a recognition by the host governments that tourism is a powerful engine of growth and a generator of foreign exchange earnings."

Purpose of Visit % of Tourists	
Leisure, Recreation and Holiday	45.45
Visiting Friends and Relatives	12.01
Business and Professional	23.00
Health and Treatment	2.20
Religion and Pilgrimage	12.24
Others	5.10
Total	100

Fig 1: Purpose Of Visit to India

(Source: Ministry Of tourism,
Government of India)

As per reports in Opportunities in Malaysian Tourism Industry (2007-2009)

<http://www.marketresearch.com/product/display.asp?productid=1806312>, the key findings include the following.

- Singapore, Thailand and Indonesia are important sources of visitors for Malaysia. Beyond ASEAN, tourist arrivals from China and India will remain an important influence throughout the forecast period (2008-2012) as the majority of Chinese tend to weigh their spending towards consumer purchases as opposed to luxury hotel accommodation.
- The promotion of Education Tourism will continue to be expanded to expedite the development of Malaysia as a preferred destination for international students. The projected foreign exchange earnings from this potential source of growth are estimated at RM 900 Million by 2010.

- It is expected that expenditure by international tourists in Malaysia will increase at a CAGR of 6.63% during the forecasted period.
- Increasing disposable income in Malaysia will open the opportunities for both outbound and domestic tourism. It is expected that per head disposable income in the country will increase at a CAGR of 5.06% during 2008-2012.
- It is expected that MICE (Meetings, Incentives, Conventions & Exhibitions) industry will be one of the major contributors to the Malaysian tourism industry.

Ministry of Tourism, Government of India in its Annual Report, has recognized the need for further promoting India through “Incredible India” campaign. The ministry aims to capitalize on the rich and varied history of India as well as it being the centre of origin of two important religions, Hinduism and Buddhism. Apart from the fact that India has a rich history, at present it is also becoming a preferred location for medical tourism as well as a recognized centre for higher education. The vast coastline of India, along with the charms of Northeast India beckons tourists who are nature lovers. The purpose of visit of tourists coming to India has been summarized in the Table 2:

Although around 50 percent of the tourists coming to India go for leisure and recreation, many

attractive destinations can be promoted apart from the beaches of Goa (a state in Western India) and Andaman and Nicobar Islands (chain of Islands in the Bay of Bengal, towards the east of India). In the following table some of the factors attracting tourists as well as facilitating them have been compared between the two countries. As can be clearly seen from the Fig 2, opportunity for attracting tourists is comparatively more in India than Malaysia. Throughout the length and breadth of Malaysia, not much cultural diversity can be witnessed. India on the other hand is known for its diversities and different cultures. India has got more than double the number of international airports, nine times more the number of heritage sites, three times more the number of beaches, four times more railway connectivity and ten times the area of Malaysia. But still the number of tourists coming to India is substantially less than the numbers coming to Malaysia as is shown in Table 3.

Parameters	<i>Malaysia</i>	<i>India</i>
No Of International Airports	5	11
Cultural Heritage Sites	3	28
No of Beaches	7	21
Railway connectivity	1,699 km	64,015 km
Uniqueness	Eco-tourism	Historical and religious
Area	328,600 sq.km	3,287,263 sq. km

Table 2: Comparison of Attractions and Infrastructural support between India and Malaysia

Year	<i>Malaysia</i> Arrivals (million)	<i>Malaysia</i> Receipts (USD) billion	<i>India</i> Arrivals (million)	<i>India</i> Receipts (USD) billion
2001	12.7	5.363	2.283	3.016
2002	13.2	7.502	2.073	3.012
2003	10.5	7.967	2.726	5.145
2004	15.7	6.572	3.457	5.588
2005	16.4	9.176	3.919	6.624
2006	17.4	9.889	4.447	7.805
2007	20.9	11.222	4.977	8.872
2008	22.1	14.322	5.287	10.146
2009	23.6	15.872	5.113	10.992

Table 3: Comparative chart of arrivals and receipts of Tourists in India and Malaysia

(Source: Ministry of tourism, Malaysia; Ministry of Tourism India)

METHODOLOGY OF DATA AND ANALYSIS

Scope of the Study

Tourism is one industry which has been actively promoted in India as well as Malaysia. India went on to promote itself as a tourist destination through the "Incredible India" campaign, which was launched at different film festivals around the globe as well as at

national level through advertisements promoting the different states of India. Malaysia on the other hand has been promoting itself as a tourist destination through the “Malaysia- Truly Asia” campaign.

Both these countries are developing countries and Asian countries. The study aims to study the difference in the purpose of visit or similarities in the tourist’s decision making process when he/she opts for this as a vacation spot. The study would also be of help to researchers and decision makers as it is both informative and instructive.

Objectives of the study

The survey was conducted with the following objectives:

- To assess demand profiles of the foreign tourists visiting the county, this may include demand for tourist places, tourism goods and services including transport and accommodation, travel agencies, cultural services, recreation and other entertainment services.

- To assess the detailed expenditure pattern of foreign tourists visiting India, and estimates of total expenditure at all India level on specific tourism products such as accommodation, food & beverage services, passenger transport, tour operators and tourist guide services, transport equipment on rental, cultural services, recreation and other entertainment services and other tourism related products and services.

- To evaluate the performance of existing tourist facilities in the country.

- To assess the number of outbound tourists to various countries.

- To undertake market segmentation analysis.

- To estimate the average duration of stay of foreign tourists in India.

- To obtain demographic, economic and social profiles of foreign tourists visiting India and the motivational factors responsible for attracting them to India.

Research Design

Descriptive: Descriptive statistics has been used to describe the basic features of the data in this study. Simple summaries about the sample and the measures have been provided. With simple graphics analysis, the quantitative data has been analyzed.

Sources of data

Secondary: Annual report of Malaysia Tourism Ministry 2006 – 2008

Annual report of India Tourism Ministry 2006 – 2008

Annual report of World Travel and Tourism Council

“Incredible India” report 2006 - Tourism Ministry of India

Primary: Questionnaire –close ended and multiple choice questions.

Research Instrument

Questionnaire –close ended and multiple choice questions were asked
Questionnaire attached as annexure

Sampling Method: Restricted random sampling

SAMPLE SIZE

A sample of 100 tourists was surveyed. The sample foreign tourists were contacted for collection of detailed information in a structured schedule at 2 exit points – Delhi and Agra.

Limitations of the study

Language Barrier: Some of the tourists, they come from South Asia countries likes Sri Lanka, Bangladesh, Nepal can speak Hindi very well but they do not know English.

Time Constraint: Researcher had only few weeks to do this survey questionnaire, one week in New Delhi and one week in Agra. The availability of foreign tourists for interview was constrained by the time at their disposal

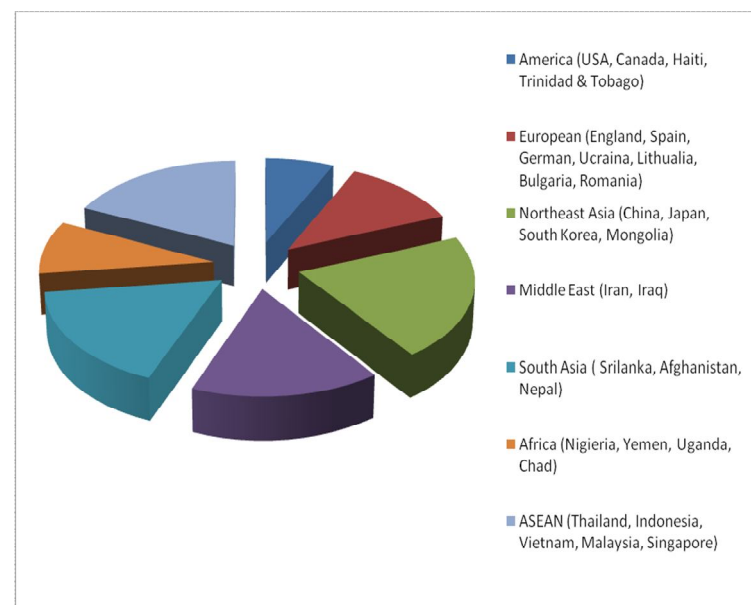
Sample Constraint.

SURVEY RESULT**I. DEMOGRAPHICS****Nationality**

An estimated 100 foreign tourists come to India from various country covered in the survey during March 2010. The major countries accounted for 39% of foreign tourist are East Asia, out of America 7%, European 12%, South

America (USA, Canada, Haiti, Trinidad & Tobago)	7%
European (England, Spain, German, Ukraine, Lithuania, Bulgaria, Romania)	12%
Northeast Asia (China, Japan, South Korea, Mongolia)	21%
Middle East (Iran, Iraq)	16%
South Asia (Sri Lanka, Afghanistan, Nepal)	17%
Africa (Nigeria, Yemen, Uganda, Chad)	9%
ASEAN (Thailand, Indonesia, Vietnam, Malaysia, Singapore)	18%
Total	100%

Asia countries 17% and Middle East 16%.

**Gender**

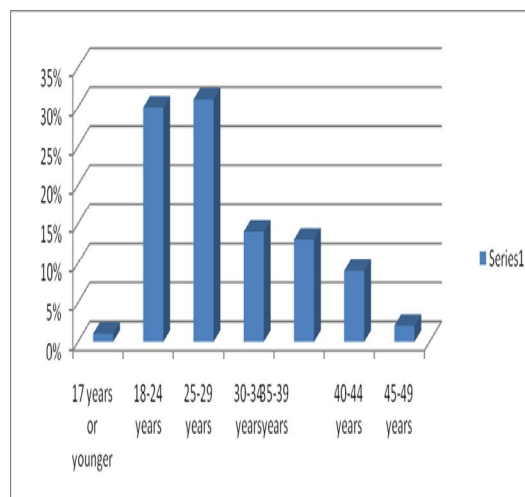
Nearly 62% of foreign tourists who visited India were males. The ratio varied from country to country.

Male	62
Female	38
Total	100

Age

The tourists were classified into seven-age groups viz., upto seventeen, eighteen to twenty four, twenty five to twenty nine, thirty to thirty four, thirty five to thirty nine, forty to forty four and forty five to forty nine. Nearly 60% of the tourists belonged to the age-group eighteen to thirty, the next highest group was thirty to thirty five (14%).

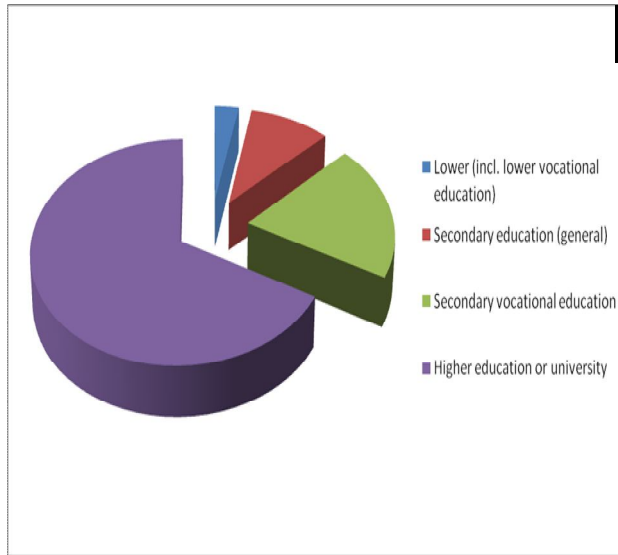
17 years or younger	1%
18-24 years	30%
25-29 years	31%
30-34 years	14%
35-39 years	13%
40-44 years	9%
45-49 years	2%
More than 50 years	0%
Total	100%



Education

The tourists were also classified on the basis of educational levels. The survey reveals that nearly 67% of the foreign nationals visiting India were graduates and postgraduates at higher education or university; only 3% tourist at lower vocational education.

Lower (incl. lower vocational education)	3%
Secondary education (general)	10%
Secondary vocational education	20%
Higher education or university	67%
Total	100%



Total	100%
-------	------

III. MOTIVATIONAL FACTORS

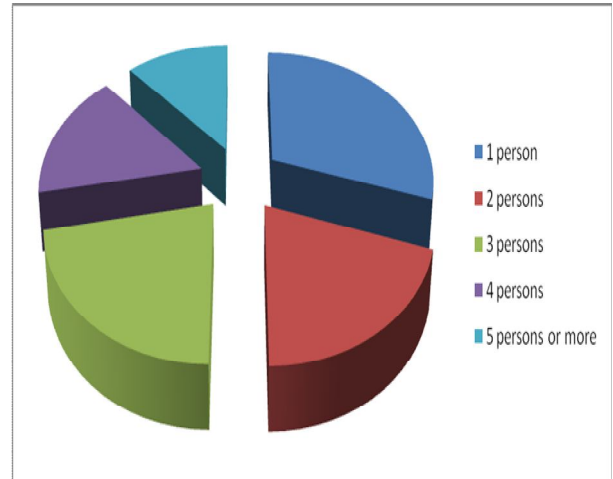
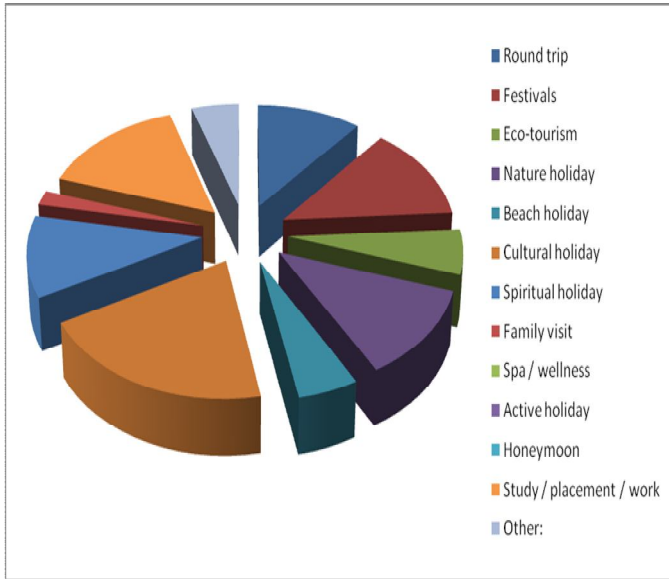
The survey shows that 67% tourists visited for leisure, recreation & holiday, 15% for business, study purpose and 11% for religion and pilgrimage, 2% visited to meet friends and relatives in India, remaining 5% visited India for other purposes.

II. SOURCE OF INFORMATION

Family/friends are major (30%) source of information, followed by Internet (22.4%), travel agents/tour operators (16%), travel magazines/films/T.V. etc. (4%) and other (13.6%).

Where did you get the information about this destination	
Travel Agency	16%
Family/friends	32%
Internet	22.40%
Papers/Magazines	4%
Travel Brochures	1.60%
TV/Radio	4%
Tourism board	0.80%
Travel guides	5.60%
Other	13.60%

What was the main purpose of your holiday to India?	
Round trip	11%
Festivals	13%
Eco-tourism	6%
Nature holiday	12%
Beach holiday	5%
Cultural holiday	20%
Spiritual holiday	11%
Family visit	2%
Spa / wellness	
Active holiday	
Honeymoon	
Study / placement / work	15%
Other:	5%
Total	100%



IV. TRAVEL PATTERN

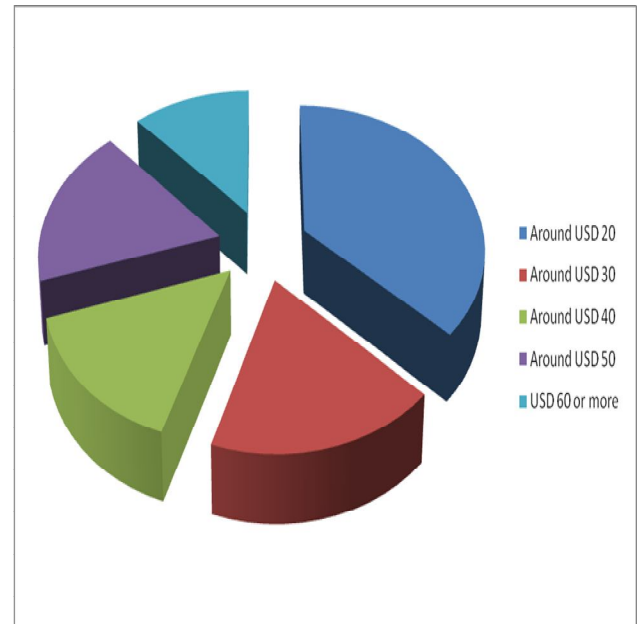
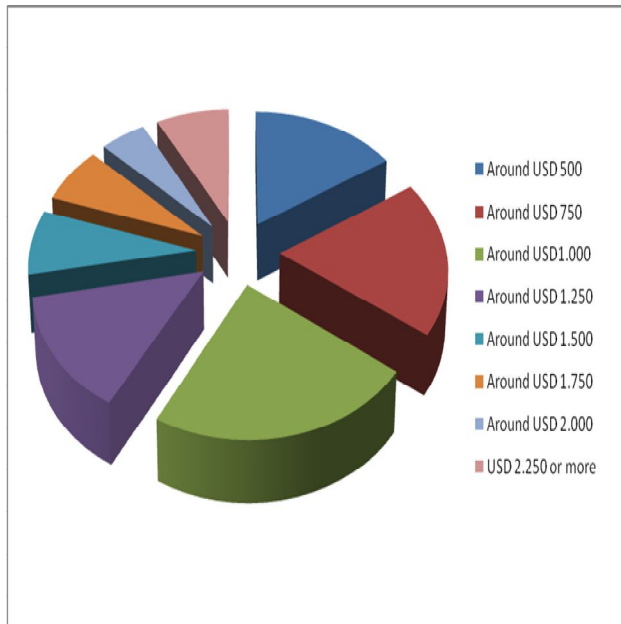
The analysis of travel pattern shows that 30% of tourists traveled alone, 20% traveled with two persons, 22% traveled in a group of 3 persons, 16% in a group of four persons, 12% in a group of five persons and more.

1 person	30%
2 persons	20%
3 persons	22%
4 persons	16%
5 persons or more	12%
Total	100%

V. EXPENDITURE PATTERN

The analysis of tourist expenditure shows that 23% tourists spent around 1000 USD, 19% tourists spent around 750 USD, 14% spent around 500 USD and only 8% tourist spent above 2250 USD.

<i>What were the travel and lodging expenses of this trip to India per person?</i>	
Around USD 500	16%
Around USD 750	19%
Around USD 1.000	23%
Around USD 1.250	14%
Around USD 1.500	8%
Around USD 1.750	7%
Around USD 2.000	5%
USD 2.250 or more	8%
Total	100%



From data survey we find that 37% of tourists spent 20 USD per day in India – excluding travel and lodging expenses and 18% of tourist spent 30 USD per day, 15% of tourists spent 40 USD, 18% of tourists spent around 50 USD and 12% of tourists spent more than 60 USD.

VI. TRANSPORT USED MOST IN INDIA

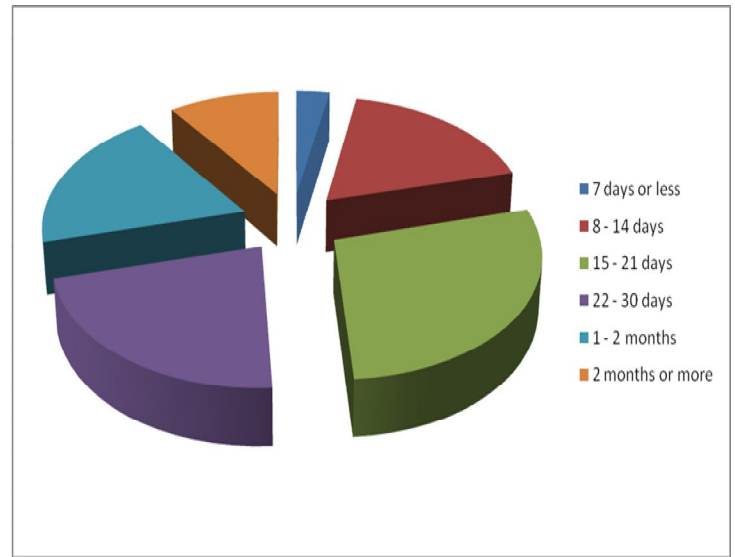
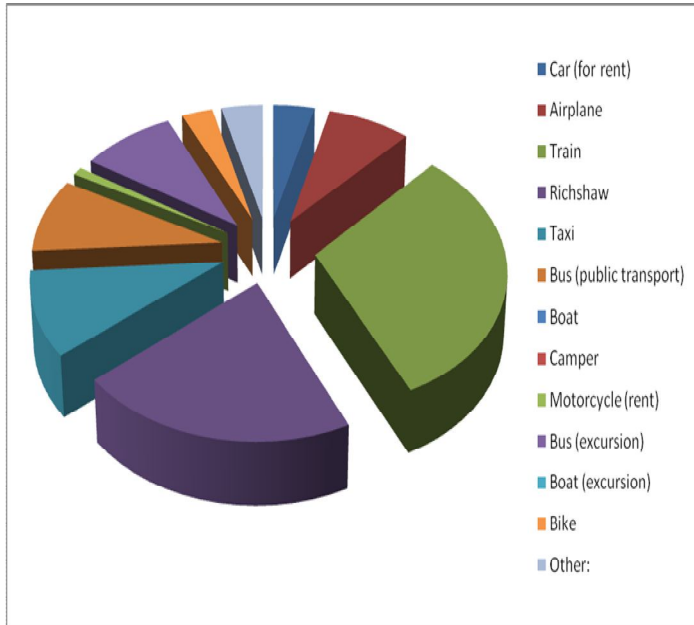
The major mode of travel within India used by foreign tourists shows that almost one-third used Rail, remaining one-third used Road and 10% used air transportation.

<i>Spend per person per day in India, excluding travel and lodging expenses</i>	
Around USD 20	37%
Around USD 30	18%
Around USD 40	15%
Around USD 50	18%
USD 60 or more	12%
Total	100%

What means of transport did you use most in India?	
Car (for rent)	4%
Airplane	8%
Train	31%
Rickshaw	21%
Taxi	10%
Bus (public transport)	9%
Motorcycle (rent)	1%

Bus (excursion)	9%
Bike	3%
Other:	4%
Total	100%

22 - 30 days	22%
1 - 2 months	19%
2 months or more	10%
Total	100%



VII. NUMBER OF DAYS STAYED

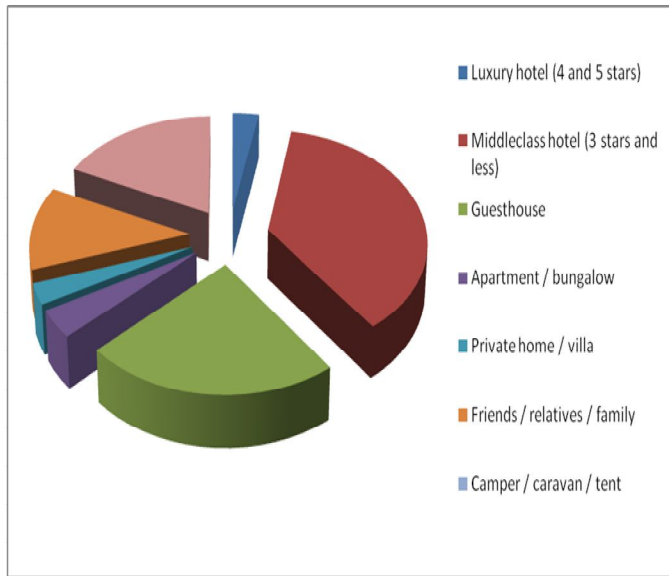
There are 68% tourists have more than one week to four weeks stayed in India, only 3% of tourists stayed one week or less. Around 30% tourists stayed more than one month to two months in India.

How long do you have holiday in India?	
7 days or less	3%
8 - 14 days	18%
15 - 21 days	28%

VIII. ACCOMMODATION

There are 60% tourists stayed in middleclass hotel; 3% spending in luxury hotel (4 and 5 stars). 12% tourists come to their friends, relatives or family.

At what kind of accommodation did you stay in India?	
Luxury hotel (4 and 5 stars)	3%
Middleclass hotel (3 stars and less)	37%
Guesthouse	23%
Apartment / bungalow	4%
Private home / villa	3%
Friends / relatives / family	12%
Other	18%
Total	100%

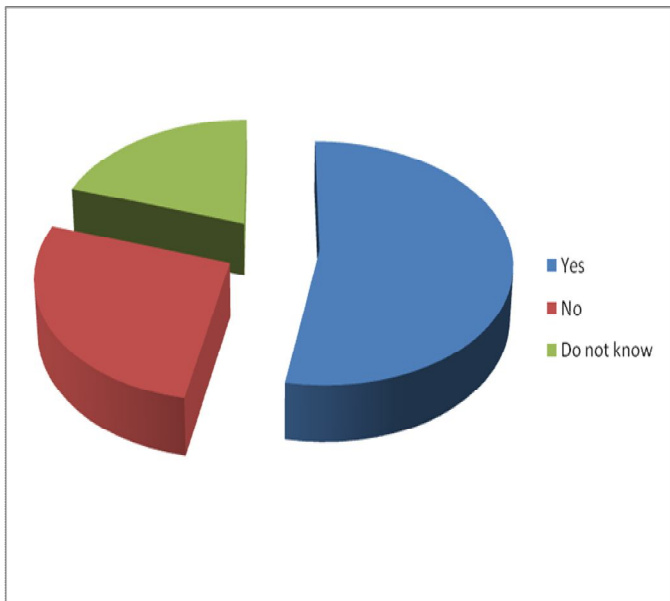
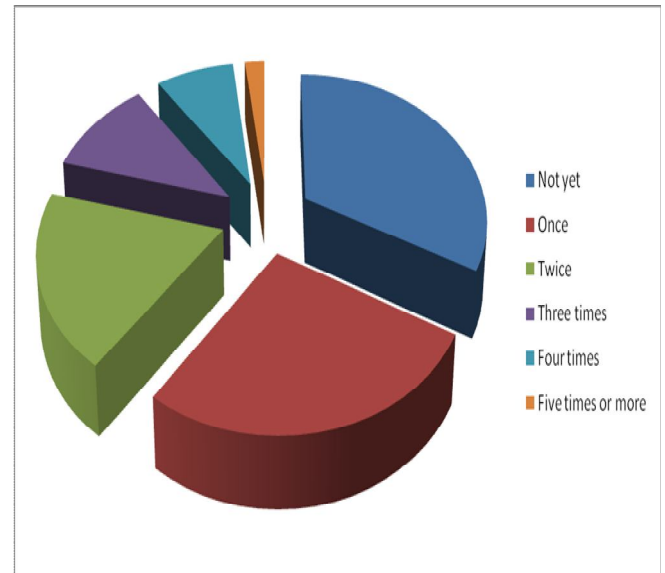


IX. REPEAT VISITS

Yes	53%
No	27%
Do Not Know	20%
Total	100%

While 33% of the tourists visited India for the first time, for 27% it was the second visit, 19% third visit and 2% of them had visited India more than five times on earlier occasions.

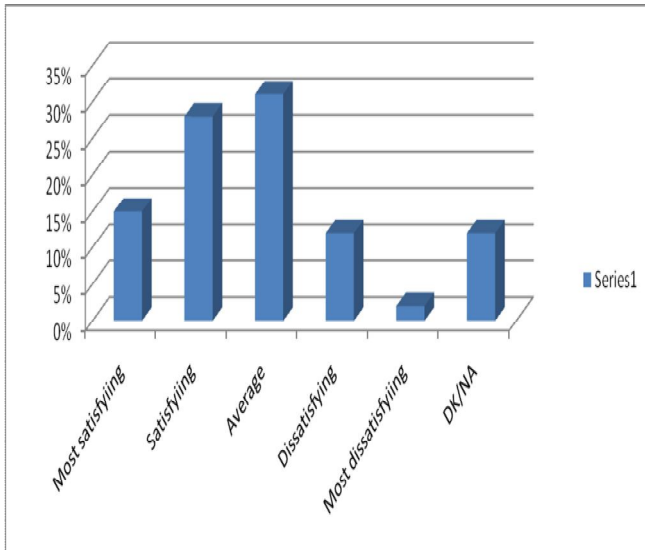
<i>How many times did you visit India before?</i>	
Not yet	33%
Once	27%
Twice	19%
Three times	11%
Four times	8%
Five times or more	2%
Total	100%



X. VALUATION OF STAY IN INDIA

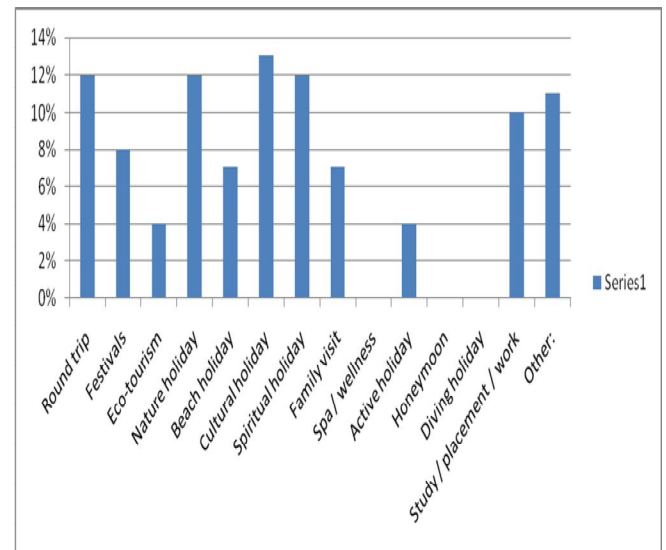
The tourists were asked to evaluate their overall satisfaction level as Most satisfying, Satisfying, Average, Disappointing and most Dissatisfying. 53% of the tourists evaluated their overall visit to India as Most Satisfying, 27% as Satisfying, 20% as Average and only 14% rated it as Disappointing and most Dissatisfying.

What is your valuation of your stay in India?	
Most satisfying	15%
Satisfying	28%
Average	31%
Dissatisfying	12%
Most dissatisfying	2%
DK/NA	12%
Total	100%



India, remaining 11% visited India for other purposes.

What would be the main purpose of your next visit to India?	
Round trip	12%
Festivals	8%
Eco-tourism	4%
Nature holiday	12%
Beach holiday	7%
Cultural holiday	13%
Spiritual holiday	12%
Family visit	7%
Spa / wellness	
Active holiday	4%
Honeymoon	
Diving holiday	
Study / placement / work	10%
Other:	11%
Total	100%



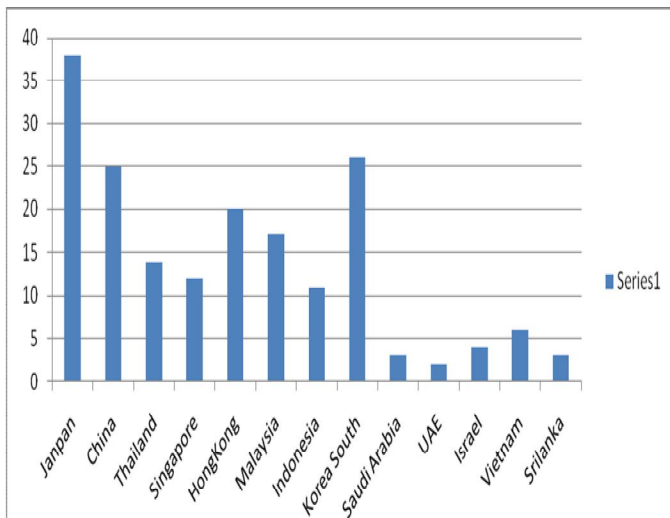
XI. MAIN PURPOSE OF NEXT VISIT TO INDIA

The survey shows that 56% tourists visited for leisure, recreation & holiday, 10% for business, study purpose and 12% for religion and pilgrimage, 7% visited to meet friends and relatives in

XII. FAVOURITE HOLIDAY DESTINATION COUNTRIES IN ASIA IN THE NEXT FEW YEARS

In total 181 ideas show out, Japan got 38 ideas, Korea South with 26, China with 25, Hong Kong had 20 and Malaysia got 17 ideas - East Asian countries are favorite holiday destinations of tourists.

<i>Favorite holiday destination countries in Asia in the next few years</i>	
Japan	38
China	25
Thailand	14
Singapore	12
Hong Kong	20
Malaysia	17
Indonesia	11
Korea South	26
Saudi Arabia	3
UAE	2
Israel	4
Vietnam	6
Srilanka	3
Total	181

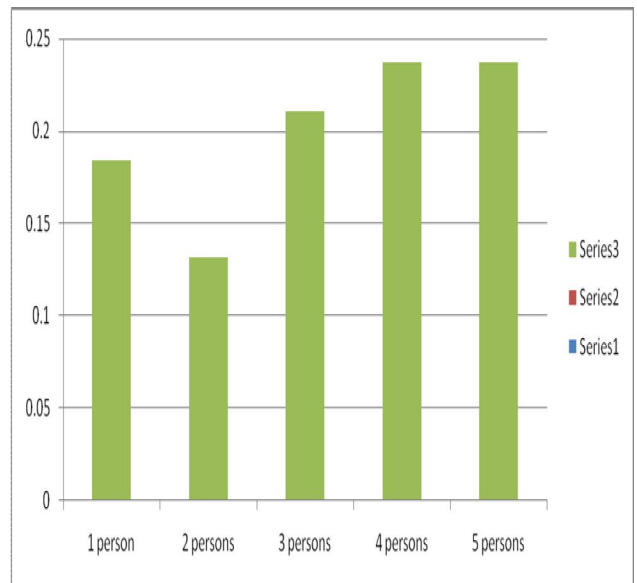


XIII. WOMAN TRAVEL PATTERN

In 38% of woman tourists to India, only 18.42% one person comes to India, more than 80% of them come to India with group of two, three, four, five persons.

We find that tourists feel not safe, harassment. The inflow of the foreign tourists could be ensured only when they are provided safe and secure environment.

Female travel pattern	
1 person	18.42%
2 persons	13.15%
3 persons	21.07%
4 persons	23.68%
5 persons	23.68%
Total	100.00%

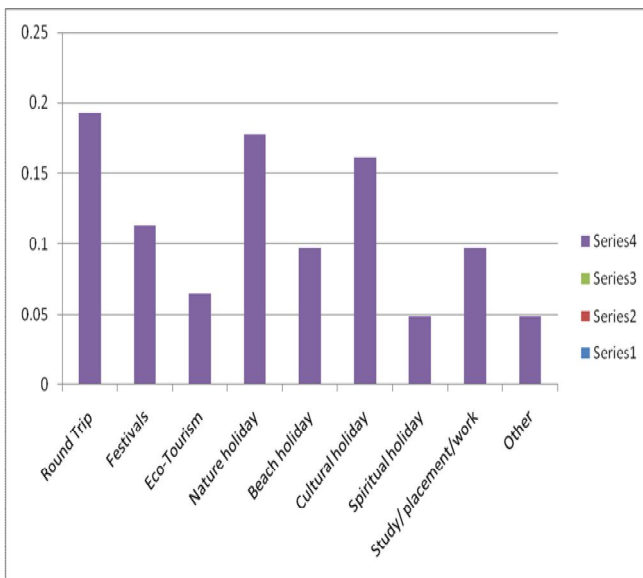


XIV. MAIN PURPOSE OF NEXT

Main purpose of next visit to India - tourists under 30 years old	
Round Trip	19.35%
Festivals	11.29%
Eco-Tourism	6.45%
Nature holiday	17.75%
Beach holiday	9.67%
Cultural holiday	16.12%
Spiritual holiday	4.85%
Study/ placement/work	9.67%
Other	4.85%
Total	100.00%

VISIT TO INDIA – TOURISTS UNDER 30 YEARS OLD

We find the trend of more than 80% tourists under 30 years old would like to come to India for leisure, recreation & holiday; 4.85% for spiritual holiday purpose; around 10% come to India with study, business purpose.



Short Term Measures

Finding: One

i. A vast majority of respondents (63%) had reported over charging, by taxi drivers, lack of manners, ill-informed tourist guides and tour operators, misbehavior with ladies, cheating, etc in many tourist destinations like Agra, Delhi...etc.

Suggestions

i. The police should be given adequate powers and additional staff to enforce discipline and change the mind set of different players in the tourism sector. State governments may examine the feasibility of constituting a special tourist police force to position at different tourist centers/ tourists spots, monuments.

ii. A code of conduct should be framed and publicized for tourist guides, taxi drivers, tour operators, travel agents etc. Concerned associations should be partnered with this proposed activity.

Finding: Two

i. About 90% tourists remarked that there were inordinate delays at immigration counters. These delays were attributable to lack of space and inadequate computerization, coupled

with heavy rush during morning hours (midnight to 3-4 A.M.)

Suggestion

i. The authorities concerned at international airports must ensure availability of more space so as to enable the Ministry of Home Affairs (MOHA) to locate additional counters and deploy more immigration officers.

ii. The paper work involved in immigration should be reduced in line with international practice i.e; all outgoing passengers as well as incoming nationals (Indian citizens) should not have to fill in an elaborate form, and incoming foreigners should also be required to fill in only simple forms that are amenable to easy and fast processing for generation of essential input for concerned authorities like Home Affairs, DOT, Civil Aviation etc.

Medium / Long term Measures

Finding: Three

i) Sizeable percentage of tourists (55 %) opined that most of the approach roads in tourist destinations were in bad conditions causing great inconvenience and delays.

Suggestion

i) A comprehensive study should be commissioned by the Ministry of Transport in collaboration with Department of Tourism to ascertain the status of approach roads and prepare a time bound action plan to rectify the defects to make them motorable and travel friendly in major tourist destinations.

Finding: Four

i) The affluent countries of Western Europe, North America, Japan and Australia are very far from India. Several foreign nationals (about 30%) cited that high cost of international travel stood in the way of attracting a larger number of tourists from these countries. The only way to reduce per capita cost of international travel is through group tours and charter flights.

Suggestion

i) Recent steps taken by the government of India by 1) Abolition of the inland air travel tax, 2) scrapping the basic fare and foreign travel tax of Rs. 500, 3) The reduction in excise duty on aviation turbine fuel to 8% from 16% etc will go a long way in bringing down the cost of travel. Similarly the state governments should relook at exorbitant luxury taxes, sales tax etc. This will substantially

reduce package tour cost thus increasing the flow of tourists. Overseas and domestic tour operators should be encouraged to organize package tours with a focus on specific tourism products like medical tourism, spiritual tourism, eco tourism etc.

ii) Efforts must be made to draw ethnic groups other than Indians from Southeast Asian countries to places in India with Buddhist relics and scenic and cultural attractions. Steps initiated recently should be intensified and spread throughout the region. Opening of new air connectivity under the Open Skies Policy and the SAARC agreements should facilitate such flights.

Finding: Five

i) The awareness of unique tourism products like Health tourism facilities are very poor. **Suggestion**

i) Developing suitable persons/ agencies to bring out directories / brochures on unique tourism products, like Health tourism, and distribute to Indian missions abroad, tour operators, travel agents etc. in the target source markets.

Finding: Six

Entry fees to monuments and heritage sites are higher for foreigners than collected from the Indian Nationals

Suggestion

i) Uniform fares for foreign and Indian nationals should be charged so that wide spread feeling of discrimination is reduced.

Finding : Seven

i) India has not succeeded in attracting family groups to visit India together in large numbers as is evident from the very low percentage of children, housewives and females amongst the tourists (38% woman). There is a great need for attracting tourists to visit India for the pure pleasure she offers. Historical sites, palaces and architectural monuments, hills, beaches and forests, wildlife, religious and folk festivals, music and dances film and theatre, handicrafts of India must be all great attractions to foreigners.

Suggestion

i) Group tourism should be encouraged by devising suitable strategies and promotional measures to attract younger generations, housewives etc. About 58% of the tourists were repeat visitors.

Finding: Eight

i) More than 30 % of the tourists visit India on their own motivation or at the influence of others who had visited India earlier.

Suggestion

i) It is important that the tourists leave India with a good impression in their minds so that they would, in turn, influence others to visit India. Factors, which irritate the tourists, like cumbersome immigration and customs procedures, unethical traders, difficulties in air or rail bookings for travel within India, unsanitary conditions at places of stay and travel must therefore be given high priority for improvement.

The Marketing Strategy should take into account the following factors: The Asian tourism market is growing at a rapid rate. This makes the case for stiff competition among different nations to attract maximum tourists; Technical assistance is required to draw comprehensive master plans and also to review and improve the existing plans; The competitive tourism trade warrants a constant and consistent marketing of new and exclusive destinations, which includes upgrading the existing ones; For a number of years in the past, the emphasis has been on marketing India as a cultural destination with rich and diverse religious history and many pilgrimage attractions. In the present scenario, there is a need to diversify the

tourism product and lay more emphasis on projecting India as a modern country for a comfortable holiday with a choice of beaches, historical attractions.

REFERENCES

1. Kotler, Philip. –Lane, Kevin. Marketing Management. (Pearson Publication's, 13th Edition)
2. Malhotra, K Naresh. Marketing Research, An Applied Orientation, (Pearson Education, 2008)
3. Lovelock, Christopher. Wirtz, Jochen .Chatterjee, Jayanta. Services Marketing.(Prentice Hall, 2002).
4. Annual report of Ministry Tourism Government of India – 2008-2009.
5. Annual report of Malaysia Tourism 2008 – 2009.
6. World Travel & Tourism council report on India and Malaysia Tourism.
7. **ASEAN Journal on Hospitality & Tourism, February 2009 - Chief Editor:** Myra P. Gunawan - Institut Teknologi Bandung, Bandung, Indonesia
8. Opportunities in Malaysian Tourism Industry (2007-2009)
<http://www.marketresearch.com/product/display.asp?productid=1806312> August 25, 2008.
- 9.http://www.researchandmarkets.com/research/a628d1/malaysia_tourism-
Authors: Rashid, Z.A., Mohamed Sharif Bashir
Author Affiliation: Department of Economics, Faculty of Economics and Management, University Putra Malaysia, PM, 43400 Serdang, Selangor, Malaysia.
10. Singh, Amrik.Outlook of Asia Pacific Tourism Industry: Current Trends and Future (Singh, Amrik is an instructor, Department of Parks, Recreation and Tourism, University of Utah, Salt Lake City, Utah - email: amrik.singh@health.utah.edu)
- 11.<http://www.tourismstat.com/survey/public/survey>.

Clinical and Computational Study of Geometry & Hemodynamics of Arterial Stenosis

Krittika Dasgupta, Abhirup Roy Choudhury, Abhijit Chanda, Debabrata Nag

Abstract— Stenosis is abnormal narrowing of blood vessels. The presence of stenosis in arteries may cause critical flow conditions. It may finally lead to stroke and heart-attack. A clinical study has been done on more than 130 patients along with computational study using 2D axisymmetric rigid model of stenosis in the carotid artery. Assumed shapes of deposition zone and the degree of occlusion used in the analysis were taken from clinical data. The Navier-Stokes equations for incompressible fluid flow have been considered as the governing equations and it has been solved with varying flow parameters using standard CFD software package. The radial velocity profiles at various points of the flow field, the centerline velocity plot and the centerline pressure plots have been obtained from computational study and compared with the clinical data.

Index Terms— Arterial flow, Clinical validation, Computational Fluid Dynamics, Hemodynamics, Mathematical Modeling, Stenosis, Stenosis geometry.

1 INTRODUCTION

IN the present century arterial stenosis is one of the major causes behind death in all parts of the globe.

Arterial Stenosis is abnormal narrowing or restriction present in the inner wall of blood vessel due to the deposition of cholesterol, fatty materials, cellular waste etc. It may happen in all large or small arteries, commonly in Coronary artery, Carotid artery, and Peripheral artery. In the present case study we only emphasize on Carotid artery stenosis. Carotid artery is one of the larger arteries, present in our neck region. The normal geometry of the artery is divided into three segments, Common Carotid Artery (CCA), External Carotid Artery (ECA) & Internal Carotid Artery (ICA). The deposition of plaque may vary in shape: Artery to complex structures and in dimension. Flow through these complex structures is commonly associated with flow separation, stagnation, recirculation, secondary vortex motion, plaque rupture etc.

Efforts have been made to model stenosis and its complex hemodynamics by Computational Fluid Dynamics (CFD) and experimental analyses since 1990's. Ku and others have made detailed studies on the fluid mechanics of vascular systems hemodynamic changes due to stenoses [1, 2]. Johnston and Kilpatrick (1991) studied the effect of geometrical irregularities in the wall of a stenosed artery for Reynolds numbers from 20 to 1000 [3]. Tang et al [1995-1998] used 3D models for steady viscous flow in an

elastic stenotic tube with various stenosis stiffness and pressure conditions [4]. In past experiments blood flow has been considered both as Newtonian or Non-Newtonian fluid depending upon the radius of the blood vessel. Hemodynamic studies have been made for both steady and pulsatile flows. However, no special emphasis is given in the stenosis geometry and previous studies used idealized models using definite curves (Cosine curve, Smooth curves, Irregular Geometry). In this study, a detail care has been taken to obtain more realistic stenosis geometry after going through more than 130 patient's real time Ultrasound Doppler Examination data.

2 CLINICAL STUDY

2.1 Data Collection

More than 130 ultrasound images of vascular stenosis have been acquired for our analysis of Carotid arterial stenosis. All these patient data have been collected randomly from different well-known multi specialty hospitals in eastern India. It has been ensured during data collection and throughout the work that no patient identity is revealed. Only the information about age and sex has been noted along with other necessary clinical information for every patient data.

2.2 Study and Analysis

Each patient data are reviewed thoroughly and very carefully to identify the common geometry and occurrence of stenosis/plaque in the artery. Maximum and minimum deposition heights, length of constriction, percentage of diametric reduction are also noted for each and every data.

- Krittika DasGupta: pursuing Masters Degree program in Bio-medical Engineering Jadavpur University, India.
E-mail: krittika.dasgupta@gmail.com
- Abhirup Roy Choudhury: pursuing Bachelors Degree program in Mechanical Engineering in Jadavpur University, India.
E-mail: abhirup1408@gmail.com

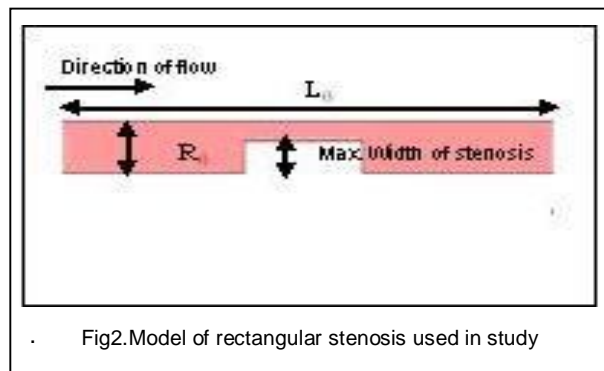
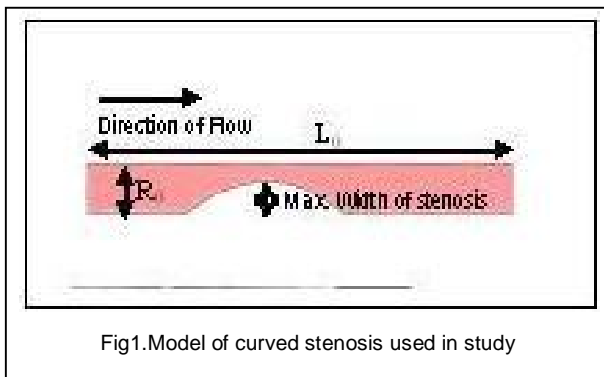
3 COMPUTATIONAL STUDY

3.1 Physical Model

In the following study, an axis-symmetric geometry has been developed by considering the carotid artery to be a long straight pipe with radius $R_0 = D_0/2$ and length $L_0 = 875R_0$, where D_0 is taken as 5.7mm (validated from different medical books).

The statistical analysis of the 130 patients reveals two dominant geometries with varied dimensions:

1. Curved shape (Fig 1)
2. Rectangular shape (Fig 2)



The occurrence of the deposition shows three dominant patterns

1. Single sided deposition
2. Axis-symmetric deposition
3. Non axis-symmetric deposition

All of them are considered with a maximum diametric constriction of 62% which can be specified as a moderate degree of stenosis, as a constriction of less than 50% is considered mild and above 70% is considered severe in most medical literature.

Sufficient length of the artery downstream of the stenosis has been taken so that the blood coming out of the constricted region is fully developed at the outlet of the artery. The upstream length for all of the stenosis is considered at $Z=0.031$.

3.2 Mathematical Model

3.2.1 Governing equations

The blood flow can be considered to be Newtonian when flowing through large arteries [1, 12]. In our study, as the common carotid artery has been dealt with the flow is considered Newtonian, laminar, steady-state and incompressible.

The incompressible Navier-Stokes Equations along with the continuity equation have been used as the governing equation for modeling the fluid flow.

$$\rho(\mathbf{u} \cdot \nabla)\mathbf{u} = \nabla \cdot [-p\mathbf{I} + \nu(\nabla\mathbf{u} + (\nabla\mathbf{u})^T)] \quad (1)$$

$$\nabla \cdot \mathbf{u} = 0 \quad (2)$$

Where u is the axial velocity, p is the axial pressure, ν is the dynamic viscosity and ρ is the density of blood and T is the transpose matrix. Equation (1) is the momentum balance equation and equation (2) is the continuity equation. In the current study the density of blood has been considered as 1050 Kg/m³ and the dynamic viscosity as 0.00345 Pa.s.

3.2.2 Boundary conditions

The imposed boundary conditions are:

1. A fully developed velocity profile at the inlet. The equation of the velocity profile is parabolic as expected in laminar flow

$$u(r) = \bar{u} \left[1 - \left(\frac{r}{R} \right)^2 \right] \quad (3)$$

where,

$u(r)$ = radial velocity at an arbitrary radius

\bar{u} = mean velocity

R = the radius of the artery

r = the radius at which the velocity is to be obtained.

2. A zero pressure with no viscous stress condition at the outlet.
3. A no-slip condition at all the walls.
i.e. $\mathbf{u} = 0$

3.3 Numerical Procedure

Standard Finite Element CFD based software COMSOL® 3.5a has been used for the solution of the problems. The solver type is parametric and the solver used is Direct (PARDISO).

3.4 Mesh Details and Grid Sensitivity Test

A free mesh consisting of triangular elements has been used in the study with the maximum possible refinement. The mesh has been refined in the vicinity of the constrict-

tions so as to present a more accurate picture of their effects on the blood flow. In the curved constriction, 14900 elements and for the rectangular constriction 15369 elements have been used for solution of the problems.

Grid sensitivity tests for all the simulations have been performed. For all of the stenosis geometries, there have been no noticeable changes in results when the grids have been refined above the values mentioned. So the above refinement of meshes is used in our subsequent studies.

3.5 Code Vallidation

In the absence of any standardized data regarding the haemodynamics of stenosed arteries, results of flow through a straight pipe without any constriction has validated the code. The plug fluid flow considered at the inlet is fully developed after a certain distance from it. The radial velocity profile of the fully developed flow is parabolic and the maximum velocity is the centerline velocity and its value is twice the mean velocity. All these results are fully compliant with the known results of classical fluid mechanics.

4 RESULTS

4.1 Clinical Results

Both of the carotid arteries have been viewed starting from CCA to ICA and ECA the sites of deposition is as follows:

TABLE1
AREA OF OCCURRENCE OF PLAQUE

Deposition of plaque along the artery	Occurrence
Common Carotid Artery (CCA)	36.15%
Internal Carotid Artery (ICA)	30.7%
External Carotid Artery (ECA)	12.3%
Bifurcation / Bulb region	20%

In this 2D longitudinal and cross sectional ultra sound images plaque is visible in either side as well as both (inner and outer) sides of the inner vessel wall. [“Outer side” means upper wall boundary of the 2D ultrasound image. Clinically it is towards periphery of the neck region and the “Inner side” is the lower wall boundary of the 2D ultra sound image. Clinically it is away from the periphery of the neck region]. This type of both-sided plaque formation shows axis symmetric and non-axis symmetric formations from single to multiple appearances.

Now, when we examine the 2D B-mode (black and white mode) longitudinal section images for both side depositions more carefully, it shows maximum 18.51% data with very small almost, non-measurable one sided deposition, where other side contributes for a good degree of diametric reduction. Calcification is prominent in 25% cases for the over all batch. Cross sectional images

indicate symmetric and asymmetric deposition.

TABLE2
APPEARANCE OF PLAQUE

Appearance	Occurrence
Single sided Deposition	54.04%
Both Sided Deposition	45.96%
Multiple Deposition	19%(Both side), 1.4% (single side)
Outer wall Deposition	14.92%
Inner wall Deposition	85.07%

Four common shapes are seen which can be broadly categorized as Cosine shaped, Rectangular shaped, conical shape and spherical or elliptical shaped. Apart from this few irregular geometries are also observed. The common trend is towards the Cosine shaped geometry (Fig.3) and rectangular geometry (Fig.4), but any possible combination of abovementioned geometry is noticed for the deposition in both walls.

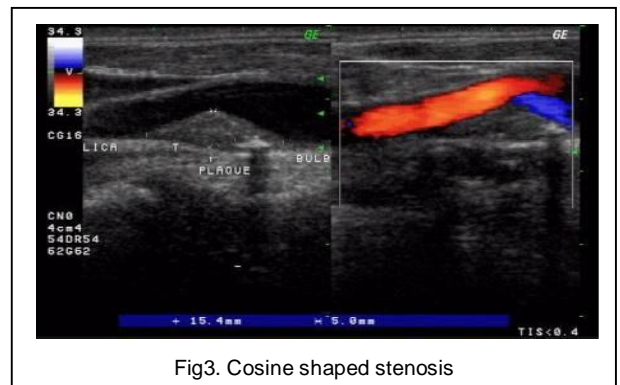


Fig3. Cosine shaped stenosis

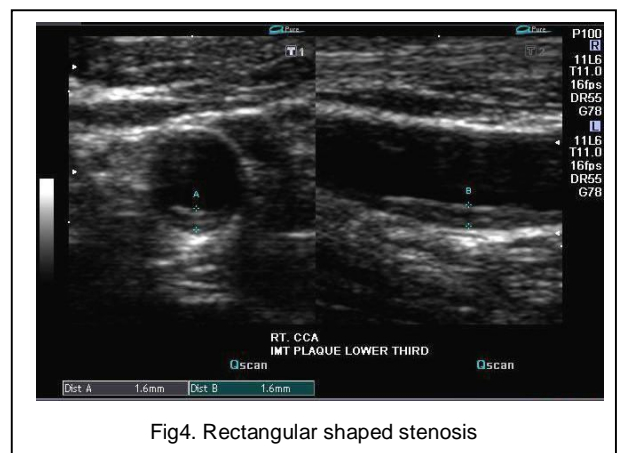


Fig4. Rectangular shaped stenosis

Size variation is very much prominent through out the batch. Large size plaques are present along with multiple small size plaques. For the both (inner and outer) sided deposition a comparison in length and width of geometry

is provided in Fig 5 and Fig 6.

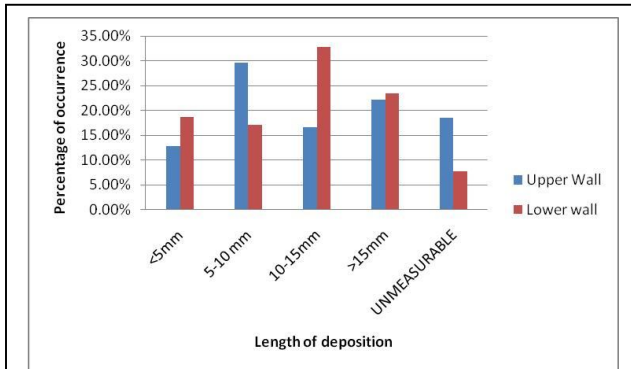


Fig5. Comparison of length of deposition in both walls

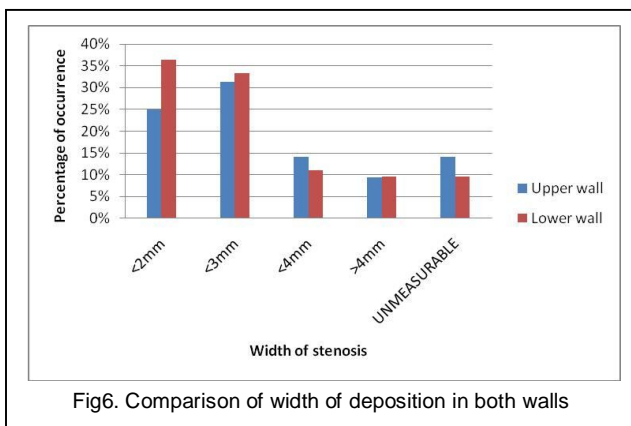


Fig6. Comparison of width of deposition in both walls

From the Fig 5, it is clear that in all the cases of small (<5mm), large (10-15mm) and very large (>15mm) deposition, inner side plaque dimension is much more than outer side. Even in case of maximum width of Deposition the combined effect of both inner and outer wall maximum width provides a high degree of diametric reduction.

More stenosis is found in Male patients as compared to Females falling in same age group, though the sample volume of the clinical study was not so large. Data shows a chance of stenosis is more for person crossing the age of 50-70 years depending on their foodhabbit, life sty and past medical records.

4.2 Computational Results

In the available literatures, blood has been found to flow with Reynolds Number (Re) between 100 to 1000. So in this study, for both the stenosis geometries, the flow has been studied with Re = 100, 400, 800 and 1000. A zone of recirculation and an irreversible pressure rise have been observed at the outlet of the stenosis for both rectangular and cosine model. The following points have been observed by studying the simulated results of the rectangular and curved stenoses.

4.2.1 Centerline pressure plot

From the centerline pressure plot, it has been observed that at the inlet of the stenosis the pressure fall is higher for higher values of Reynolds Number (Re). Even negative pressure values have been found in case of Re = 1000. But at the outlet of the stenosis, the flows with higher Re show higher values of pressure. So, the irreversible pressure rise increases with increasing values of Re. (Fig. 7)

When comparing the pressure profiles of rectangular and curved stenoses at a fixed Reynolds Number, the irreversible pressure rise has been found to be higher in case of the rectangular geometry. At Re = 1000, the pressure rise for the rectangular stenosis is found to be 23% higher than the curved geometry. (Fig. 8)

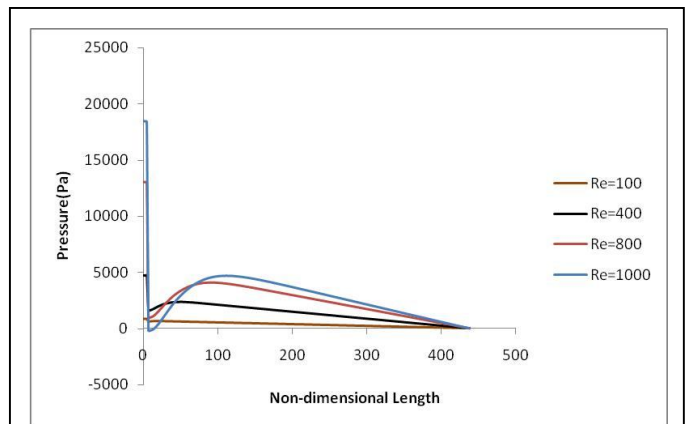


Fig7. The centerline pressure plots of a curved stenosis at different Reynolds Numbers

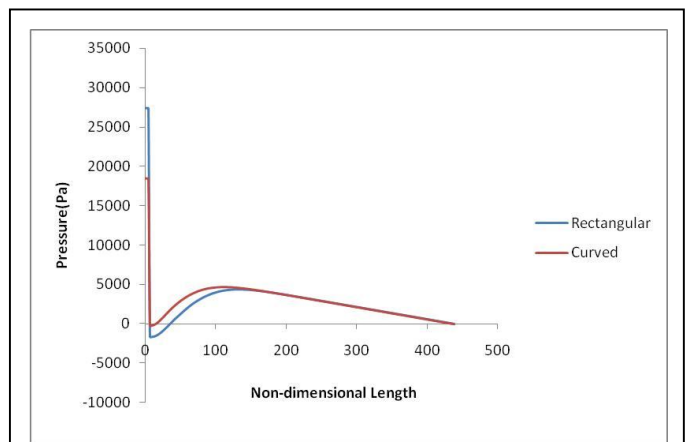


Fig8. Centerline pressure plots of Rectangular and curved stenosis at Reynolds Number=1000

4.2.2 Radial velocity field

As seen in Fig.9 the radial velocity field at the outlet of the stenosis shows negative velocity and the maximum value of the negative velocity is higher for higher values of Re. (For the rectangular stenosis, at the stenosis inlet, the maximum velocity has been found to shift from the centerline).

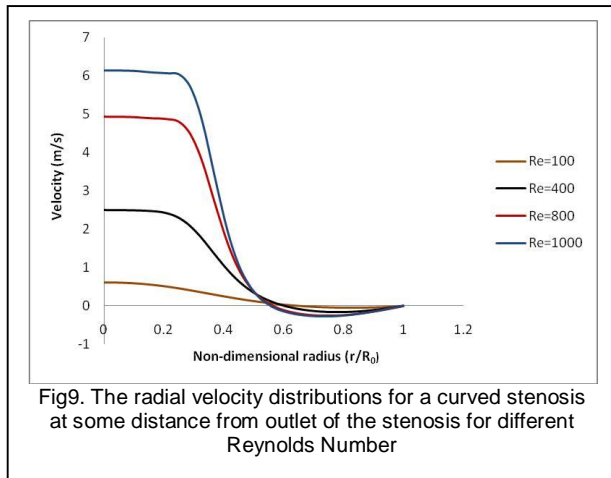


Fig9. The radial velocity distributions for a curved stenosis at some distance from outlet of the stenosis for different Reynolds Number

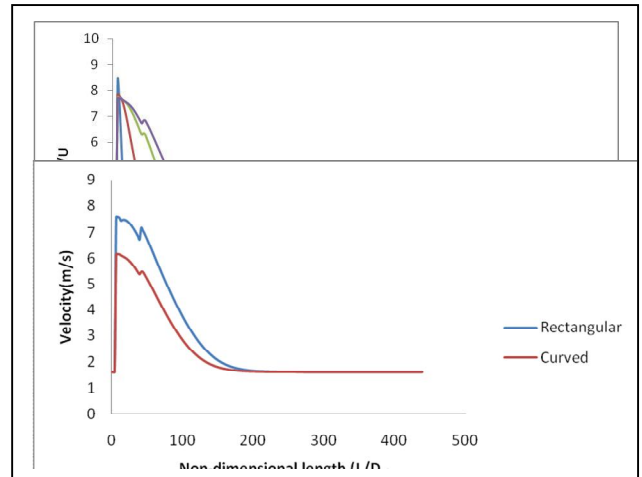


Fig12. The centerline velocity plots of rectangular and curved stenoses at Reynolds Number=1000

The maximum velocity in case of rectangular profile is higher than the curved geometry.

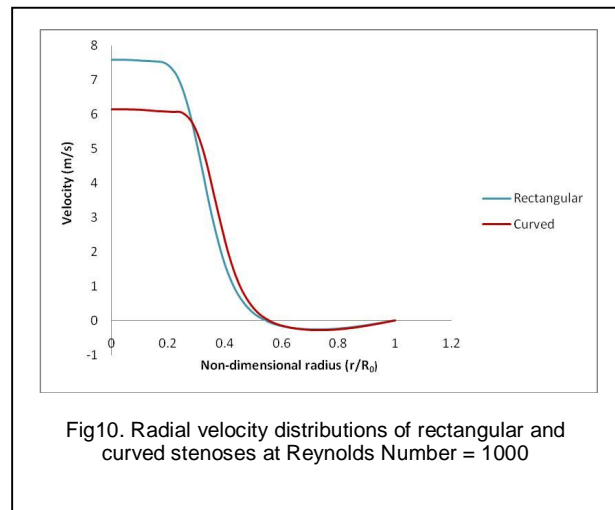
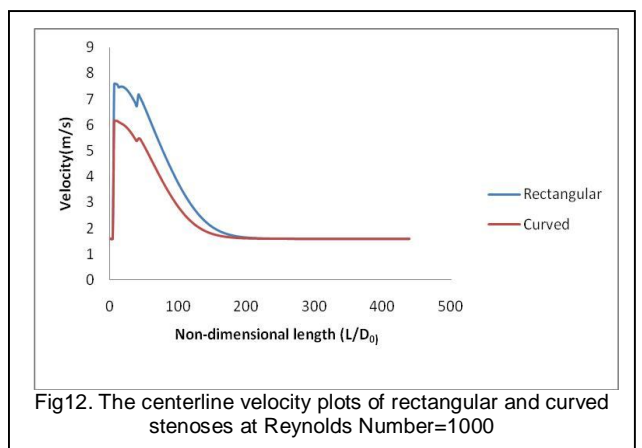


Fig10. Radial velocity distributions of rectangular and curved stenoses at Reynolds Number = 1000



In our current study we are concentrating on two main single sided deposition geometry i.e. cosine shape and rectangular shape. But in our clinical study we not only get single deposition we encountered multiple depositions near about 20% cases.

4.2.3 Centerline velocity field

From the centerline velocity plot (Fig 11), the maximum velocity in the entire sub-domain has been found in a zone near the outlet of the stenosis for all the values of Re used. Also the length of reattachment of the flow after the flow separation has been found to increase with increasing values of Reynolds Number.

The reattachment length of the rectangular stenosis is around 10% more than the curved stenosis, as can be concluded from (Fig. 12).

4.2.3 Centerline velocity field

The recirculation lengths of both the rectangular and curved stenoses plotted against the respective Reynolds Numbers show an almost linear variation. From this graph (Fig 13) it can be seen very clearly that the recirculation lengths of the rectangular stenosis is higher than a curved stenosis for the same value of Reynolds number.

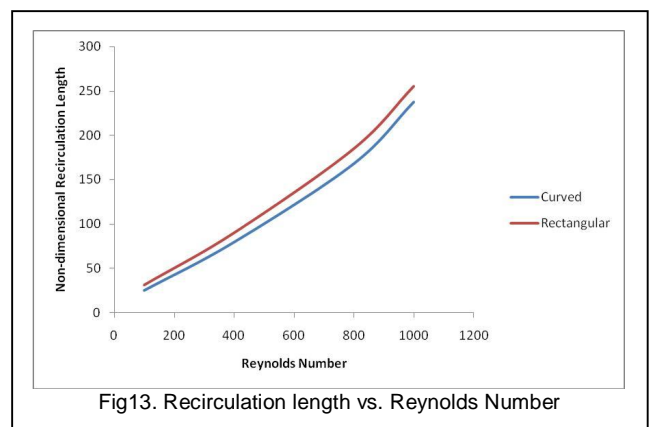


Fig13. Recirculation length vs. Reynolds Number

The data set also reflects good amount of large sized plaque accompanied by smaller one or multiple depositions, where their dimension, placement, and appearance varies. Table 3 shows the distance between two adjacent

depositions, where almost no gap is considered for distance of < 1mm, small gap is considered for the distance between 1mm – 3mm, large gap is taken for the distance between 3mm – 6mm, and above this all the distances are considered as very large gap.

TABLE3
DISTANCE BETWEEN MULTIPLE STENOSIS

Distance Between Multiple Deposition	Percentage of occurrence
Almost no gap between two adjacent formation	16.66%
Small gap between two adjacent formation	50%
Large gap between two adjacent deposition	25%
Very Large gap between two adjacent deposition	8.33%

From the clinical study it is found that multiple stenoses often occur in reality. It causes much more complexity in flow pattern. With such multiple depositions, size and shape of recirculation zone is expected to suffer variation. Numerical Modeling with such real life situations involving multiple stenosis is being done currently and will be reported later

5 INFERENCE

From the clinical study of more than 130 patients with carotid artery stenosis we come to the common inferences after detailed study of it, which are stated as follows.

More deposition is seen in ICA with respect to ECA proves the haemodynamic nature and construction of bifurcation geometry of human Carotid artery. A good amount of deposition is noted in the bulb and near bifurcation zone conforms the fact of flow separation and stagnation and deposition of substances in this type of regions.

Lower wall deposition is maximum for both single and both sided deposition stating stagnation of particles in the lower surface of a flowing liquid for a pipe flow.

From the geometry and appearance of multiple plaque and its internal distance it is clear that a single deposition initiate another deposition in its locality an along time this depositions merges to form a more complex and large structure which is clinic ally more dangerous.

In the present work the flow of blood through stenosed arteries has been studied by considering blood to be a Newtonian fluid and the flow to be laminar by varying the Reynolds Number.

For fixed stenosis geometry, as the pressure rise increases with increasing Reynolds Number, the heart has to supply even more pressure to overcome this adverse pressure gradient. Thus the effort of the heart increases, leading to angina (pain in the heart). Also as the recirculation zone is higher for higher values of the flow velocity, the tendency of the stenosis to propagate increases. This is because the stenosis aggravates with higher lengths of

re-attachment.

From the comparison of the rectangular and curved stenosis, it is inferred that as the extent of the recirculation length is longer in case of the rectangular stenosis than the curved one, the rectangular stenosis has a higher tendency to propagate. This is expected from the shapes of the stenoses. The rectangular stenosis presents itself as a bluff body in the line of flow of blood while the curved stenosis is more streamlined. Also that the irreversible pressure rise for the rectangular geometry is higher proves that it is more severe than a curved stenosis of the same maximum constriction due to causes already discussed.

It can also be predicted from the above study that as depositions continue to occur downstream of a curved stenosis, it will eventually develop into a rectangular stenosis if enough time is available. So, the adverse effects of a stenosis essentially increase with time.

ACKNOWLEDGMENTS

A great support has been provided by Dr. Manoranjan Mahapatra, Kalinga Hospital, Bhubaneswar; Dr. Ashok Moulik, and CMRI Hospital, Kolkata by providing clinical ultrasound Doppler data and also PURSE Project, DST, Govt of India for funding the project.

REFERENCES

- [1] D.N Ku., Blood flow in arteries. Ann. Rev. Fluid Mech.vol. 29, pp. 399-434, 1997.
- [2] D.M Wootton., D.N.Ku., Fluid mechanics of vascular systems, diseases, and thrombosis. Annu. Rev. Biomed. Eng. Vol.01, pp. 299-329, 1999.
- [3] P.R. Johnston and D. Kilpatrick, Mathematical modeling of flow through an irregular arterial stenosis. Journal of Biomechanics vol. 24, pp. 1069-1077, 1991.
- [4] H.I. Anderson, R. Halden, T. Glomsaker, Effects of surface irregularities on flow resistance in differently shaped arterial stenosis. Journal of Biomechanics vol. 33, pp. 1257-1262, 2000.
- [5] D. Tang, C.Yang, D.N Ku., A 3-D thin-wall model with fluid-structure interaction for blood flow in carotid artery with symmetric and asymmetric stenosis. Computers and Structures vol. 72, pp.357-377, 1999.
- [6] C. Bertolotti, V. Deplano, Three-dimensional numerical simulation of flow through stenosed coronary bypass. Journal of Biomechanics vol. 33, pp. 1011-1022. 1999.
- [7] P.K.Mandal, An unsteady analysis of Non-Newtonian blood flow through tapered arteries with stenosis. International Journal of Non-Linear Mechanics vol. 40, pp. 151-164, 2005.
- [8] A. Yakhot., L. Grinberg, N.Nikitin, Modeling rough stenoses by an immersed-boundary method. Journal of Biomechanics vol. 38, pp. 1115-1127, 2005.
- [9] C.A Taylor., J.D Humphrey., Open problems in computational vascular biomechanics: Haemodynamics and arterial wall stenosis. Comput. Methods Appl. Mech. Engrg. Vol.198, 3514-3523, 2009.
- [10] COMSOL Multiphysics User Guide.
- [11] COMSOL Multiphysics Modelling Guide.
- [12] Jay D. Humphrey, Sherry L. Delange, An Introduction to Biomchanics.

Segmentation Techniques for Iris Recognition System

Surjeet Singh, Kulbir Singh

Abstract— A biometric system provides automatic identification of an individual based on a unique feature or characteristic possessed by the individual. Iris recognition is regarded as the most reliable and accurate biometric identification system available. Iris recognition systems capture an image of an individual's eye, the iris in the image is then segmented and normalized for feature extraction process. The performance of iris recognition systems highly depends on segmentation and normalization. This paper discusses the performance of segmentation techniques for iris recognition systems to increase the overall accuracy.

Index Terms—Active contour, Biometrics, Daugman's method, Hough Transform, Iris, Level Set method, Segmentation.

1. INTRODUCTION

Reliable personal recognition is critical to many processes. Nowadays, modern societies give higher relevance to systems that contribute to the increase of security and reliability, essentially due to terrorism and other extremism or illegal acts. In this context, the use of biometric systems has been increasingly encouraged by public and private entities in order to replace or improve traditional security systems. Basically, the aim is to establish an identity based on who the person is, rather than on what the person possesses or what the person remembers.

Biometrics can be regarded as the automated measurement and enumeration of biological characteristics, in order to obtain a plausible quantitative value that, with high confidence, can distinguish between individuals.

Although less automatized, biometrics has been used - at least - for centuries. In the 14th century, the Portuguese writer João de Barros reported its first known application. He wrote that Chinese merchants stamped children's palm print and footprints on paper with identification purposes. In the western world, until the late 1800s the automatic recognition of individuals was largely done using "photographic memory". In 1883, the French police and anthropologist Alphonse Bertillon developed an anthropometric

system, known as Bertillonage, to fix the problem of identification of convicted criminals.

In 1880, the British scientific journal Nature published an article by Henry Faulds and William James describing the uniqueness and permanence of fingerprints. This motivated the design of the first elementary fingerprint recognition system by Sir Francis Galton and improved by Sir Edward R. Henry. Having quickly disseminated, the first fingerprint system in the United States was inaugurated by the New York State Prison Department in 1903 and the first known convicted due to fingerprint evidences was reported in 1911.

Presently, due to increasing concerns associated with security and the war on terrorism, biometrics has considerably increased its relevance. It has moved from a single and almost standardized trait (fingerprint) to the use of more than ten distinct traits.

According to Matyas Jr. and Riha [1], every biometric system depends on the features, whether genotypic or phenotypic it is based on. Similarly to Daugman [2], authors divide the biometric traits into two types. Fried [3] and A. Bromba [4] classified the origin of the biometric traits into three different types: genotypic, behavioral, and randotypic.

Following the proposal of Jain et al. [5], biometric systems can be evaluated regarding seven parameters: uniqueness, universality, permanence, collectability, performance, acceptability and circumvention.

Figure 2 contains a comparison between the most common biometric traits. Each value was obtained through averaging and weighting of the classifications proposed in [6], [7], [4], [8], [9], [10] and [11].

For the purposes of our work, one of the most important features is the ability to perform

- Surjeet Singh is currently faculty Member at Faculty of Science & Technology, The iCFAI University, Dehradun, Inida. E-mail: surjeet_singh1986@ymail.com
- Kulbir Singh is currently Assistant Professor, in Electronics & Communication Engineering Department, Thapar University, Patiala, India. E-mail: ksingh@thapar.edu

covert recognition, which can be performed by the fingerprint, face, iris and palmprint. Among these, iris must be enhanced, as it provides higher uniqueness and circumvention values.

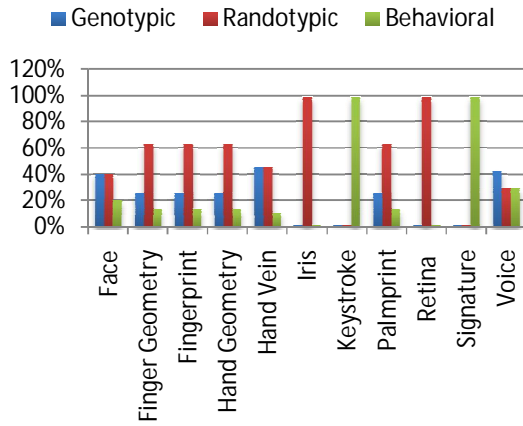


Fig 1: Factors of influence of the biometric traits.

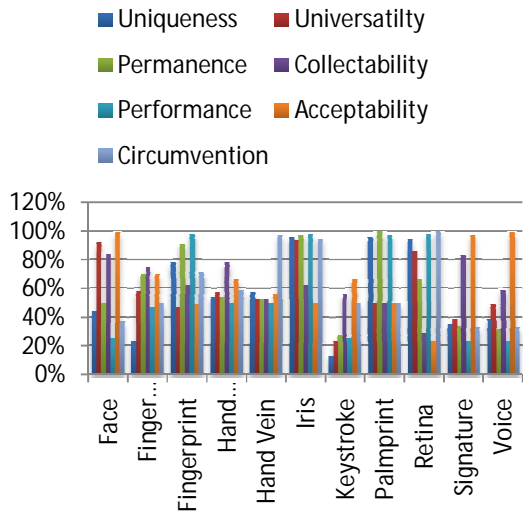


Fig 2: Comparison between the most common biometric traits (adapted and averaged from [6], [7], [4], [8], [9], [10] and [11]).

2. IRIS ANATOMY

The iris is a thin, circular structure located anterior to the lens, often compared to a diaphragm of an optical system. The centre aperture, the pupil, actually is located slightly nasal and inferior to the iris centre. Pupil size regulates retinal illumination. The diameter can vary from 1 mm to 9 mm depending on lighting conditions. The pupil is very small (miotic) in brightly lit conditions and fairly large (mydriatic) in dim illumination. The average diameter of the iris is 12 mm, and its thickness varies. It is thickest in the region of the collarette, a circular ridge approximately 1.5 mm from the pupillary margin. This slightly raised jagged ridge was the attachment site for the fetal pupillary membrane

during embryologic development. The collarette divides the iris into the pupillary zone, which encircles the pupil, and the ciliary zone, which extends from the collarette to the iris root. The colour of these two zones often differs [12].

The pupillary margin of the iris rests on the anterior surface of the lens and, in profile, the iris has a truncated cone shape such that the pupillary margin lies anterior to its peripheral termination, the iris root. The root, approximately 0.5 mm thick, is the thinnest part of the iris and joins the iris to the anterior aspect of the ciliary body. The iris divides the anterior segment of the globe into anterior and posterior chambers, and the pupil allows the aqueous humor to flow from the posterior into the anterior chamber with no resistance.

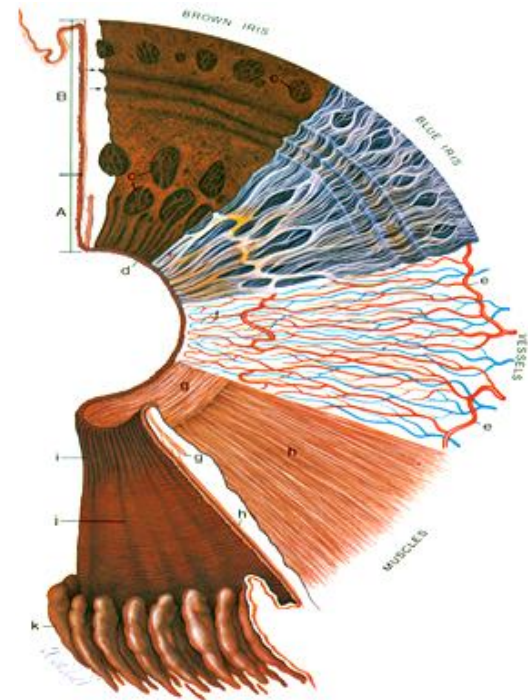


Fig 3: Surfaces and layers of the iris [12].

In figure 3 the iris cross-section shows the pupillary (A) and ciliary portions (B), and the surface view shows a brown iris with its dense, matted anterior border layer. Circular contraction furrows are shown (arrows) in the ciliary portion of the iris. Fuchs' crypts (c) are seen at either side of the collarette in the pupillary and ciliary portion and peripherally near the iris root. The pigment ruff is seen at the pupillary edge (d). The blue iris surface shows a less dense anterior border layer and more prominent trabeculae. The iris vessels are shown beginning at the major arterial circle in the ciliary body (e). Radial branches of the arteries and veins extend toward the pupillary region. The arteries form the incomplete minor arterial circle

(f), from which branches extend toward the pupil, forming capillary arcades. The sector below it demonstrates the circular arrangement of the sphincter muscle (g) and the radial processes of the dilator muscle (h). The posterior surface of the iris shows the radial contraction furrows (i) and the structural folds of Schwalbe (j). Circular contraction folds also are present in the ciliary portion. The pars plicata of the ciliary body is at (k). [13]

3. IRIS SEGMENTATION

In this paper CASIA iris image database has been used for the analysis of different segmentation algorithms. CASIA iris image database (version 1.0) includes 756 iris images from 108 eyes, hence 108 classes.

In 1993, J. Daugman [14] presented one of the most relevant methods, constituting the basis of the majority of the functioning systems. Regarding the segmentation stage, this author introduced an integrodifferential operator to find both the iris inner and outer borders. This operator remains actual and was proposed in 2004 with minor differences by Nishino and Nayar [15].

Similarly, Camus and Wildes [16] and Martin-Roche et al. [17] proposed integrodifferential operators that search the \mathbb{N}^3 space, with the objective of maximizing the equations that identify the iris borders.

Wildes [18] proposed iris segmentation through a gradient based binary edge-map construction followed by circular Hough transform. This is the most common method, that has been proposed with minor variants by Cui et al. [19], Huang et al.[20], Kong and Zhang[21], Ma et al.[22], [23] and [24].

Liam et al. [25] proposed one interesting method essentially due to its simplicity. This method is based in thresholds and in the maximization of a simple function, in order to obtain two ring parameters that correspond to iris inner and outer borders.

Du et al. [26] proposed the iris detection method based on the prior pupil segmentation. The image is further transformed into polar coordinates and the iris outer border is detected as the largest horizontal edge resultant from Sobel filtering. However, this approach may fail in case of non-concentric iris and pupil, as well as for very dark iris textures.

Morphologic operators were applied by Mira and Mayer [27] to obtain iris borders. They detected the pupillary and scleric borders by applying thresholding, image opening and closing.

Based on the assumption that the pixels' intensity of the captured image can be well represented by a mixture of three Gaussian distributions, Kim et al. [28] proposed the use of Expectation Maximization [29] algorithm to estimate the respective distribution parameters. They expected that 'Dark', 'Intermediate' and 'Bright' distributions contain the pixels corresponding to the pupil, iris and reflections areas.

3.1 Daugman's Method

This is by far the most cited method [14] in the iris recognition literature. The author assumes both pupil and iris with circular form and applies the following integrodifferential operator:

$$\max_{r,x_0,y_0} \left| G_\sigma(r) * \frac{\delta}{\delta r} \oint_{r,x_0,y_0} \frac{I(x,y)}{2\pi r} ds \right| \quad (1)$$

This operator searches over the image domain (x,y) for the maximum in the blurred (by a Gaussian Kernel $G_\sigma(r)$) partial derivative with respect to increasing radius r , of the normalized contour integral of $I(x,y)$ along a circular arc ds of radius r and center coordinates (x_0,y_0) . In other words, this method searches in the \mathbb{N}^3 space for the circumference center and radius with highest derivative values comparing to circumferences of neighbour radius.

At first the blurring factor σ is set for a coarse scale of analysis so that only the very pronounced circular transition from iris to (white) sclera is detected. Then after this strong circular boundary is more precisely estimated, a second search begins within the confined central interior of the located iris for the fainter pupillary boundary, using a finer convolution scale σ and a smaller search range defining the paths (x_0,y_0,r) contour integration. In the initial search for the outer bounds of the iris, the angular arc of contour integration ds is restricted in range to two opposing 90° cones centered on the horizontal meridian, since eyelids generally obscure the upper and lower limbus of the iris. Then in the subsequent interior search for the pupillary boundary, the arc of contour integration ds in operator (1) is restricted to the upper 270° in order to avoid the corneal specular reflection that is usually superimposed in the lower 90° cone of the iris from the illuminator located below the video camera. Taking the absolute value in (1) is not required when the operator is used first to locate the outer boundary of the iris, since the sclera is always lighter than the iris and so the smoothed partial derivative with increasing radius near the limbus is always positive. However, the pupil is not always darker than the iris, as in persons with normal early cataract or significant back-

scattered light from the lens and vitreous humor; applying the absolute value in makes the operator a good circular edge-finder regardless of such polarity-reversing conditions. With automatically tailored to the stage of search for both the pupil and limbus, and by making it correspondingly finer in successive iterations, the operator defined in has proven to be virtually infallible in locating the visible inner and outer annular boundaries of irises.

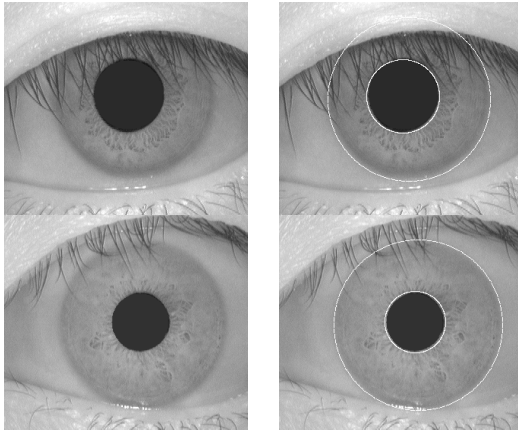


Fig 4: Segmentation by Daugman's Method

For rapid discrete implementation of the integrodifferential operator in , it is more efficient to interchange the order of convolution and differentiation and to concatenate them, before computing the discrete convolution of the resulting operator with the discrete series of undersampled sums of pixels along circular contours of increasing radius. Using the finite difference approximation to the derivative for a discrete series in ,

$$\frac{\partial}{\partial r} f(r, \theta) \approx \frac{f(r + \Delta r, \theta) - f(r - \Delta r, \theta)}{2\Delta r}$$

where Δr is a small increment in radius, and replacing the convolution and contour integrals with sums, we can derive through these manipulations an efficient discrete operator (3) for finding the inner and outer boundaries of an iris where $\Delta \theta$ is the angular sampling interval along the circular arcs, over which the summed pixel intensities represent the contour integrals expressed in .

3.2 Hough Transform

Hough transform is a standard image analysis tool for finding curves that can be defined in a parametrical form such as lines, polynomials and circles. The recognition of a global pattern is achieved using the local patterns. For instance, recognition of a circle can be achieved by considering the strong edges in an image as the local patterns and searching for the maximum value of a circular Hough transform.

Wildes et al. [18], Kong and Zhang [21], Tisse et al. [30] and Ma et al. [22] use Hough transform to localize irises. The localization method, similar to Daugman's method, is also based on the first derivative of the image. In the proposed method by Wildes, an edge map of the image is first obtained by thresholding the magnitude of the image intensity gradient:

Where G_x and G_y are the horizontal and vertical gradients of the image intensity, respectively, and G is the magnitude of the gradient. G is a Gaussian smoothing function with scaling parameter σ to select the proper scale of edge analysis. The edge map is then used in a voting process to maximize the defined Hough transform for the desired contour. Considering the obtained edge points as (x_i, y_i) , a Hough transform can be written as:

where

The limbus and pupil are both modeled as circles and the parametric function is defined as:

Assuming a circle with the center (x_c, y_c) and radius r , the edge points that are located over the circle result in a zero value of the function. The value of $f(r, \theta)$ is then transformed to $H(r, \theta)$ by the function, which represents the local pattern of the contour. The local patterns are then used in a voting procedure using the Hough transform, $H(r, \theta)$, in order to locate the proper pupil and limbus boundaries. In order to detect limbus, only vertical edge information is used. The upper and lower parts, which have the horizontal edge information, are usually covered by the two eyelids. The horizontal edge information is used for detecting the upper and lower eyelids, which are modeled as parabolic arcs.

We implemented this method in MATLAB® by first employing Canny edge detection to generate an edge map. Gradients were biased in the vertical direction for the outer iris/sclera boundary, as suggested by Wildes et al. [24]. Vertical and horizontal gradients were weighted equally for the inner iris/pupil boundary.

The range of radius values to search for was set manually, depending on the database used. For the CASIA database, values of the iris radius range from 90 to 150 pixels, while the pupil radius ranges from 28 to 75 pixels. In order to make the circle detection process more efficient and accurate, the Hough transform for the

iris/sclera boundary was performed first, then the Hough transform for the iris/pupil boundary was performed within the iris region, instead of the whole eye region, since the pupil is always within the iris region.

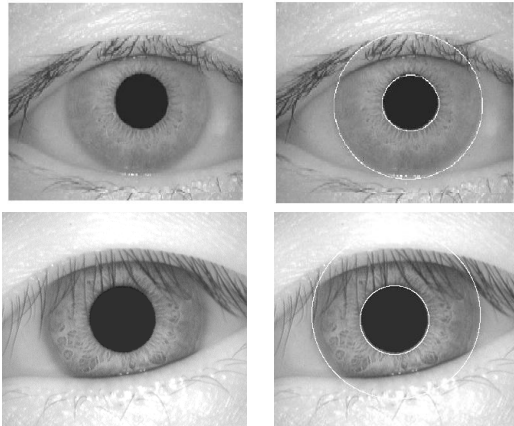


Fig 5: Segmentation by Hough Transform.

3.3 Variational Level Set Formulation of Active Contours without Re-initialization

In image segmentation, active contours are dynamic curves that moves toward the object boundaries. To achieve this goal, we explicitly define an external energy that can move the zero level curve toward the object boundaries. Let I be an image, and E be the edge indicator function defined

$$E(x, y) = \begin{cases} 1 & \text{if } |I_x| > \tau \text{ or } |I_y| > \tau \\ 0 & \text{otherwise} \end{cases}$$

where G is the Gaussian kernel with standard deviation σ . We define an external energy for a function ϕ as below

where α and β are constants, and the terms $\int \phi$ and $\int \phi^2$ are defined by

respectively, where δ is the univariate Dirac function, and H is the Heaviside function. Now, we define the following total energy functional

The external energy $\int \phi$ drives the zero level set toward the object boundaries, while the internal energy $\int \phi^2$ penalizes the deviation of ϕ from a signed distance function during its evolution. To understand the geometric meaning of the energy, we suppose that the zero level set of ϕ can be represented by a differentiable parameterized curve C . It is well known that the energy functional $\int \phi$ computes the length of the zero level curve of ϕ in the conformal metric g . The energy functional $\int \phi^2$ in g is introduced to speed

up curve evolution. Note that, when the function ϕ is constant, the energy functional in g is the area of the region Ω . The energy functional in g can be viewed as the weighted area of Ω . The coefficient of ϕ can be positive or negative, depending on the relative position of the initial contour to the object of interest. For example, if the initial contours are placed outside the object, the coefficient in the weighted area term should take positive value, so that the contours can shrink faster. If the initial contours are placed inside the object, the coefficient should take negative value to speed up the expansion of the contours.

By calculus of variations, the Gateaux derivative (first variation) of the functional in g can be written as

$$\delta J(\phi) = \int \left(-\Delta \phi + \alpha \phi - \beta \phi^3 \right) \delta \phi$$

where Δ is the Laplacian operator. Therefore, the function ϕ that minimizes this functional satisfies the Euler-Lagrange equation $-\Delta \phi + \alpha \phi - \beta \phi^3 = 0$. The steepest descent process for minimization of the functional $J(\phi)$ is the following gradient flow:

$$\frac{\partial \phi}{\partial t} = -\left(-\Delta \phi + \alpha \phi - \beta \phi^3 \right)$$

this gradient flow is the evolution equation of the level set function in the proposed method.

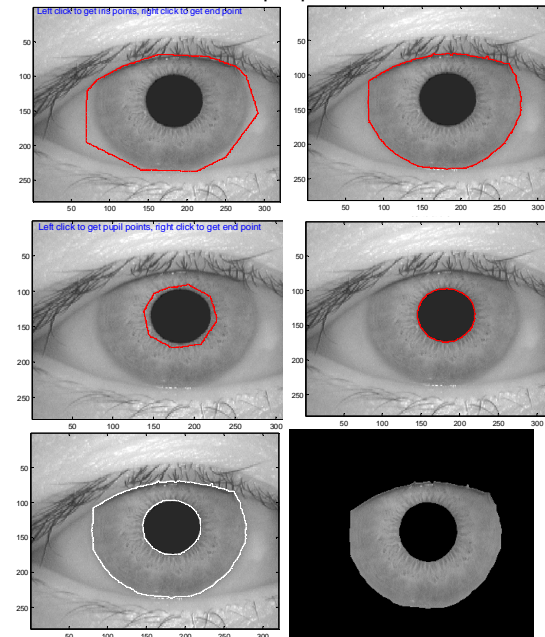


Fig 6: (a) Selected iris mask (b) iris/sclera boundary (c) selected pupil mask (d) iris/pupil boundary (e) segmented image (f) iris region.

The second and the third term in the right hand side of correspond to the gradient flows of the energy functional and , respectively, and are responsible of driving the zero level curve towards the object boundaries. To explain the effect of the first term, which is associated to the internal energy , we notice that the gradient flow

$$\frac{\partial I}{\partial t} = \text{div}(\nabla I)$$

has the factor — as diffusion rate. If , the diffusion rate is positive and the effect of this term is the usual diffusion, i.e. making more even and therefore reduce the gradient . If , the term has effect of reverse diffusion and therefore increase the gradient [31].

4. RESULTS

The automatic segmentation model using Integrodifferential equations and Hough transform proved to be successful. The CASIA database provided good segmentation, since those eye images had been taken specifically for iris recognition research and boundaries of iris pupil and sclera were clearly distinguished. For the CASIA database, the Hough transform based segmentation technique managed to correctly segment the iris region from 658 out of 756 eye images, which corresponds to a success rate of around 87% as compared to the Hough transform based segmentation technique that managed to correctly segment the iris region from 624 out of 756 eye images, which corresponds to a success rate of around 83%.

Using Integrodifferential equations and Hough transform methods on locating the pupil and limbus assume that the boundaries are perfect circles. Although the approaches are different, all these methods consider pupil and limbus as circular curves. It has been noticed that the circular assumption of the contours can lead to inappropriate boundary detection Figure 7 and 8.

The above methods of segmentation resulted in false detection due to noises such as strong boundaries of upper and lower eyelids. The strong eyelid boundaries and presence of eyelashes affected the limbus localization significantly.

We also implemented eyelashes and eyelids detection for the above two methods. The eyelid detection system proved quite successful, and managed to isolate most occluding eyelid regions. One problem was that it would sometimes isolate too much of the iris region,

which could make the recognition process less accurate, since there is less iris information. However, this is preferred over including too much of the iris region, if there is a high chance it would also include undetected eyelash and eyelid regions.

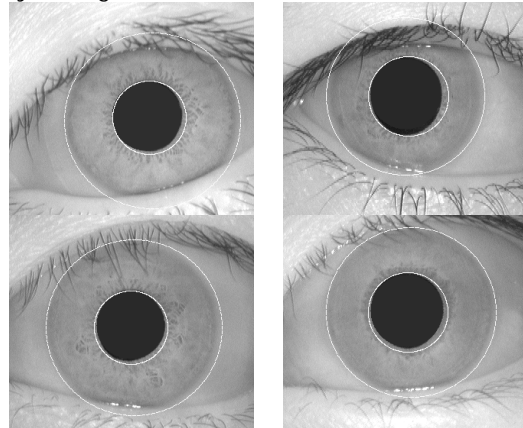


Fig 7: Illustrate the results of the Integrodifferential operator over the pupils that are not perfect circles. The circular contour does not detect pupil boundaries accurately.

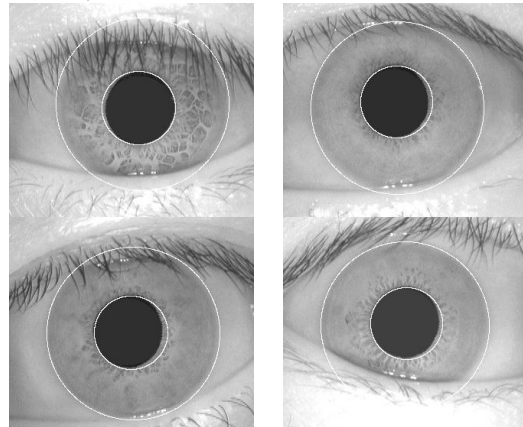


Fig 8: Illustrate the results of the Hough transform operator over the pupils that are not perfect circles. The circular contour does not detect pupil boundaries accurately.

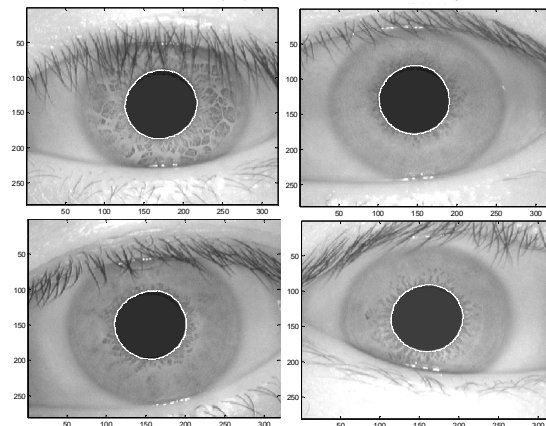


Fig 9: Illustrate the results of the active contour segmentation method based on Level set evolution without re-initialization over the pupils that are not perfect circles.

The eyelash detection system implemented for the CASIA database also proved to be successful in isolating most of the eyelashes occurring within the iris region as shown in Figure 11. A slight problem was that areas where the eyelashes were light, such as at the tips were not detected. However, these undetected areas were small when compared with the size of the iris region.

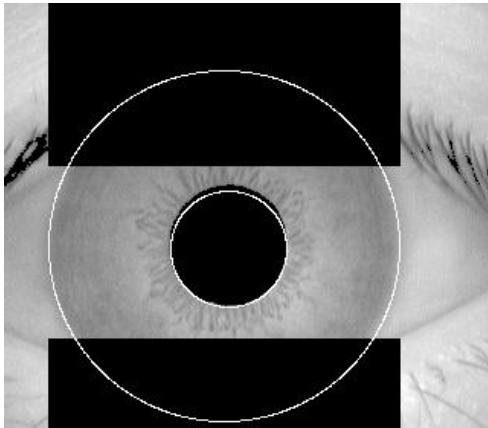


Fig 10: Automatic segmentation of image from CASIA database. Black region denote detected eyelid.

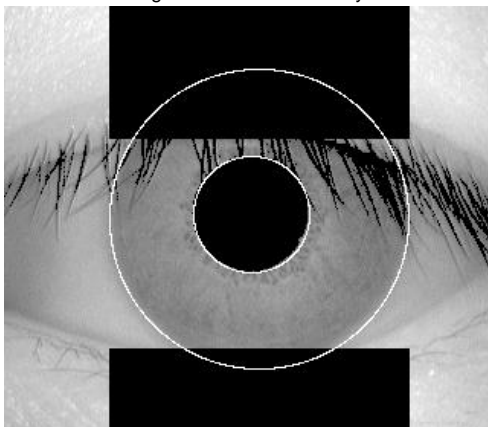


Fig 11: The eyelash detection technique, eyelash regions are detected using thresholding and denoted as black.

TABLE I
COMPARISON OF DIFFERENT SEGMENTATION TECHNIQUES.

Method	No. Of eye images	Properly Segmented	Accuracy
Daugman's Method	756	658	87%
Hough Transform	756	624	83%
Proposed Method	756	750	Approx: 100%

The proposed method of active contour segmentation based on Level set evolution without re-initialization provided perfect

segmentation results for the pupil and limbus boundaries with success rate of almost 100%. Only problem with this system was that the initial contour was to be defined for each eye image manually.

REFERENCES

- [1] V. Matyas and Z. Riha, "Toward reliable user authentication through biometrics," IEEE Security and Privacy, vol. 1, no. 3, pp. 45-49, 2003.
- [2] J. G. Daugman, "Phenotypic versus genotypic approaches to face recognition," Face Recognition: From Theory to Applications, pp. 108-123. Heidelberg: Springer-Verlag, 1998.
- [3] S. D. Fried, "Domain access control systems and methodology," http://www.itu.dk/courses/SIAS/E2005/AU22_40_01.pdf, 2004.
- [4] M. Bromba, Biometrics FAQ's, <http://www.bromba.com/faq/biofaq.htm>, 2010.
- [5] A. K. Jain, R. Bolle, and S. Pankanti, Personal Identification in networked society, 2nd edition. Kluwer Academic Publisher, E.U.A., 1999.
- [6] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 4-19, January 2004.
- [7] S. Liu and M. Silverman, "A practical guide to biometric security technology," IT Professional, vol. 3, no 1, pp. 27-32, January 2001.
- [8] Biometrics and the courts, <http://ctl.ncsc.dni.us/biomet%20web/BMIndex.html>, 2010.
- [9] Idesia's Biometric Technologies. Biometric comparison table, http://www.idesia-biometrics.com/technology/biometric_comparison_table.html, 2010.
- [10] International Biometric Group, "Which is the best biometric technology?," http://www.biometricgroup.com/reports/public/report_s/best_biometric.html, 2010.
- [11] J. D. Woodward, K. W. Webb, E. M. Newton, M. A. Bradley, D. Rubenson, K. Larson, J. Lilly, K. Smythe, B. Houghton, H. A. Pincus, J. Schachter, and P. Steinberg, "Army Biometric Applications - Identifying and Addressing Socio-Cultural Concerns," Rand Corporation, Santa Monica, 2001.
- [12] A. K. Khurana, Comprehensive Ophthalmology, New Age International (P) Ltd., 4th edition, 2007.
- [13] L. A. Remington, Clinical Anatomy of the Visual System, Elsevier Inc., 2nd edition, 2005.
- [14] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 11, pp. 1148-1161, November 1993.
- [15] K. Nishino and S. K. Nayar, "Eyes for relighting," ACM Trans. Graph., vol 23, no. 3, pp. 704-711, 2004.
- [16] T.A. Camus and R. Wildes, "Reliable and fast eye finding in close-up images," Proceedings of the IEEE 16th

- International Conference on Pattern Recognition, pp. 389–394, Quebec, August 2002.
- [17] D. Martin-Roche, C. Sanchez-Avila, and R. Sanchez-Reillo, "Iris recognition for biometric identification using dyadic wavelet transform zero-crossing," *IEEE Aerospace and Electronic Systems Magazine*, Mag. 17, no. 10, pp. 3–6, 2002.
- [18] R. P. Wildes, "Iris recognition: an emerging biometric technology," *Proceedings of the IEEE*, vol. 85, no.9, pp. 1348–1363, U.S.A., September 1997.
- [19] J. Cui, Y. Wang, T. Tan, L. Ma, and Z. Sun, "A fast and robust iris localization method based on texture segmentation," *Proceedings of the SPIE Defense and Security Symposium*, vol. 5404, pp. 401–408, August 2004.
- [20] J. Huang, Y. Wang, T. Tan, and J. Cui, "A new iris segmentation method for recognition," *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, vol. 3, pp. 23–26, 2004.
- [21] W. K. Kong and D. Zhang, "Accurate iris segmentation method based on novel reflection and eyelash detection model," *Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing*, pp. 263–266, Hong Kong, May 2001.
- [22] L. Ma, Y. Wang, and T. Tan, "Iris recognition using circular symmetric filters," *Proceedings of the 25th International Conference on Pattern Recognition*, vol. 2, pp. 414–417, Quebec, August 2002.
- [23] L. Ma, T. Tan, Y. Wang, and D. Zhang, "Personal identification based on iris texture analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 2519–2533, December 2003.
- [24] L. Ma, Y. Wang, and D. Zhang, "Efficient iris recognition by characterizing key local variations," *IEEE Transactions on Image Processing*, vol. 13, no. 6, pp. 739–750, June 2004.
- [25] L. Liam, A. Chekima, L. Fan, and J. Dargham, "Iris recognition using self organizing neural network," *Proceedings of the IEEE Student Conference on Research and Developing Systems*, pp. 169–172, Malaysia, June 2002.
- [26] Y. Du, R. Ives, D. Etter, T. Welch, and C. Chang, "A new approach to iris pattern recognition," *Proceedings of the SPIE European Symposium on Optics/Photonics in Defence and Security*, vol. 5612, pp. 104–116, October 2004.
- [27] J. Mira and J. Mayer, "Image feature extraction for application of biometric identification of iris - a morphological approach," *Proceedings of the 16th Brazilian Symposium on Computer Graphics and Image Processing*, pp. 391–398, Brazil, October 2003.
- [28] J. Kim, S. Cho, and J. Choi, "Iris recognition using wavelet features," *Kluwer Academic Publishers, Journal of VLSI Signal Processing*, no. 38, pp. 147–256, November 2004.
- [29] A. P. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistic Society*, vol. 39, pp. 1–38, 1977.
- [30] C. Tisse, L. Martin, L. Torres, and M. Robert, "Person identification technique using human iris recognition," *Proceedings of the 25th International Conference on Vision Interface*, pp. 294–299, Calgary, July 2002.
- [31] C. Li, C. Xu, C. Gui, and M. D. Fox, "Level Set Evolution Without Re-initialization: A New Variational Formulation," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 430–436, 2005.

Investigation of X-ray Plasmon Satellites in Rare Earth Compounds

Ajay Vikram Singh, Dr. Sameer Sinha

Abstract- We have Investigate and study to X-ray Plasmon Satellites in Rare earth compounds (La_2CuO_4 , Nd_2CuO_4 , Gd_2CuO_4 , PrNiSb_2 , NdNiSb_2 , $\text{Pr}(\text{OH})_3$, $\text{Nd}(\text{OH})_3$, $\text{Sm}(\text{OH})_3$)

Keywords- Surface Plasmon Satellites, Relative Intensity & Energy Separation

INTRODUCTION

IN the characteristic X-ray Spectra, Diagram as well as non Diagram lines are present. Those lines which fit in the conventional energy level diagram are called Diagram lines. & those lines which do not fit in the conventional energy level diagram are called non diagram lines. It is also known as “Satellites or Second order lines”. Satellites are generally of weak intensity lines & are found close to more intense parent line. The satellites which are observed on higher energy side are called high energy satellites (HES) whereas those are observed on lower energy side are called lower energy satellites (LES). First Siegbahn & Stenstroem observed these satellites in the K-Spectra of element from Cr (24) to Ge (32) while Coster Theraeus & Richtmyer in the L-Spectra of element from Cu (29) to Sb (51) & Hajlmar, Hindberg & Hirsch in the M-Spectra of elements from Yb (70) to U (92). Several theories were proposed from time to time to explain the origin of these satellites. Out of these theories the plasmon theory is found to be the most suitable theory especially for those satellites.

Plasmon theory was first proposed by Bohm & Pines which are extended by Houston, Ferrel, Noziers & Pines. According to this theory the low energy plasmon satellites are emitted when valence electron excites a plasmon during the annihilation of core hole conversely if Plasmon pre exists, its energy add up to the energy of diagram line.

The radiation less reorganization of electronic shell of an atom is known as Auger effect. Auger satellites have also been observed by Korbar and Mehlhorn [1] Haynes et al. [2] Edward and Rudd [3]. Theoretical explanation for K series Auger spectrum was given by Burhop and Asaad [4] using intermediate coupling. Later on more refined theory, using relativistic and configuration interaction has been used by Listengarter [5] and Asaad [6]

In Auger primary spectra, one can also observe secondary electron peaks close to the primary peaks are produced by incident electrons which have undergone well energy losses. The most common

source of such energy loss in the excitation of collective plasma oscillations of the electrons in the solid. This gives rise to a series of plasma peaks of decreasing magnitude spaced by energy $\hbar\omega_p$ where ω_p is the frequency of plasma oscillation.

Auger peaks are also broadened by small energy losses suffered by the escaping electrons. This gives rise to a satellite on the low energy of the Auger peak. Energy loss peaks have well defined energy with to primary energy.

The involvement of Plasmon oscillation in the X-ray emission or absorption spectra of solids has been widely studied during the last few decades and has been recognized that the electron-electron interaction has played an important role.

This Paper is devoted to **Investigate and study to X-ray Plasmon Satellites in Rare earth compounds (La_2CuO_4 , Nd_2CuO_4 , Gd_2CuO_4 , PrNiSb_2 , NdNiSb_2 , $\text{Pr}(\text{OH})_3$, $\text{Nd}(\text{OH})_3$, $\text{Sm}(\text{OH})_3$)**

According to Plasmon theory, if the valence electron, before filling the core vacancy, also excites a Plasmon, then the energy $\hbar\omega_p$ needed for the excitation of Plasmon oscillation is taken from the transiting valence electron so that the emitted radiation will be derived off an energy $\hbar\omega_p$ and a low energy satellites will emitted whose separation from the main X-ray line will correspond to $\hbar\omega_p$. On the other hand if the Plasmon pre exists, during the X-ray emission process, then, on its decay it can give its energy to the transiting valence electron before it annihilates the core vacancy. Thus the energy of emitted X-ray photon will be higher than the main emission line and by an amount $\hbar\omega_p$ giving rise to high energy satellite.

MATHEMATICAL CALCULATION –

In order to confirm the involvement of Plasmon in the emission of X-ray satellites the relative intensity of single Plasmon satellites must be calculated. In this process first we deal with mathematical details of canonical transformation carried out over the model Hamiltonian of the system. Thus the energy separation ΔE of the low and high energy Plasmon satellite from the

corresponding main line should be equal to the quantum of Plasmon energy $\hbar\omega_p$ which is given by [10]

$$\Delta E = \hbar\omega_p = 28.8 \sqrt{\left(\frac{Z\sigma}{w}\right)} \text{ eV} \quad 1$$

Where Z = No. of unpaired electrons, σ = Specific gravity & w = Molecular Weight

This equation can be derived as given below .

From the classical consideration, we get the frequency of Plasmon oscillation as

$$\omega_p = \left(\frac{4\pi n e^2}{m}\right)^{1/2} \quad 2$$

Hence the amount of energy given to Plasmon becomes

$$E_p = \hbar\omega_p = h \left(\frac{4\pi n e^2}{m}\right)^{1/2}$$

In this equation we can write $n = \frac{L\sigma Z}{w}$

Where σ , Z and w are defined above and L is the Avogadro number. By putting the numerical value of constant, we get the Plasmon energy as

$$\Delta E = \hbar\omega_p = 28.8 \sqrt{\left(\frac{Z\sigma}{w}\right)} \text{ eV} \quad 3$$

Our calculated values of ΔE have been compared with the Scrocco's experimental value. And We have also calculated the relative intensity of plasmon satellites, which is different in different processes. If the excitation of plasmon occurs during the transport of the electron through the solid, it is known as extrinsic process of plasmon excitation. The plasmon can also be excited by another method known as intrinsic process. In this process, excitation of plasmon takes place simultaneously with creation of a hole. Bradshaw et al have further divided core hole excitation into two classes,

1 - Where the number of slow electrons are conserved.

2 - Where the number of slow electrons are not conserved

The Author has calculated relative intensity in both the cases with new modification in the light of Bradshaw [12] and Lengreth [13] work, which explains that not only intrinsic process but extrinsic process and their relative contribution may also contribute in relative intensities. The combined effect of intrinsic and extrinsic plasmon excitation intensity variation was suggested by Lengreth as:

$$i = \frac{I_s}{I_m} = \alpha^n \sum_{m=0}^n \frac{\left(\frac{\beta}{\alpha}\right)^m}{m!} \quad 4$$

The value of β is taken as $\beta = 0.12r_s$ which is purely intrinsic, $r_s = (47.11 / \hbar w_s)^{2/3}$ is dimensionless

parameter and $\alpha = 0.47 r_s^{1/2}$ in the place of $\alpha = (1+L)^{-1}$ used by Pardee et. al. (14). The equation (3) contains a series of terms. The first term of the equation is purely extrinsic, while second term is purely intrinsic. The other terms are containing the relative contributions of both extrinsic and intrinsic. The specialty of this formula is that each term alone or simultaneously with other terms is able to give the relative intensity. This formula also includes both the categories mentioned by Bradshaw and gives better results as compared than traditional methods for calculation of the relative intensity. Using the values of α , β and r_s in equation (4)

Using the equation (4), the author has for the first time, calculated the relative intensity of Rare earth compounds (La_2CuO_4 , Nd_2CuO_4 , Gd_2CuO_4 , PrNiSb_2 , NdNiSb_2 , Pr(OH)_3 , Nd(OH)_3 , Sm(OH)_3) and Our calculated and estimated values are in agreement with the calculated values of J. C. Parlebas et al. [15] and A. Szytula, B. Penc, A. Jezierski [16]

Reference

1. Korbar H. & Mehlhorn W.A. ; Phys. 191, (1966) 217.
2. Haynes S.K. & Velinsky, M & Velinsky L.J. ; Nucl. Phys. A99 (1967), 537.
3. Rudd M.E. & Edward & Volz, D.J. ; Phys Rev. 151, (1966), 28.
4. Asaad, W.N. & Burhop E.H.S. ; Proc. Phys. Soc., London 71, (1958), 369.
5. Listengarten, M.A. ; Bull Acad. Sci. U.S.S.R., Phys. Ser. 26 (1962), 182.
6. Asaad, W.N. ; Nucl. Phys. 66, (1965b), 494.
7. M.Scrocco in photoemission spectra of Pb.(II) halide; Phys. Rev. B25 (1982) 1535-1540 .
8. M.Scrocco , Satellites in X-ray Photo electron spectroscopy of insulator I 32 (1985) 1301-1306
9. M.Scrocco , Satellite in X-ray Photo electron spectroscopy of insulators II 32 (1985) 1307-1310
10. L.Marton , L.B.Lader and H. Mendlowitz; Adv. Electronic and Electro Physics; edited by L.M arton Academic , New York 7 (1955) , 225 .
11. Surendra poonia and S.N.Soni , Indian journal of pure and applied physics , vol.45, feb.2007 pp-119-126
12. A. M. Bradshaw, Cederbaurn S.L, Domeke W. & Krause Jour. Phys C: Solid State Phys. 7, 4503, 1974
13. D. C. Lengreth, Phys. Rev. Letter, 26, 1229, 1971
14. W. J. Pardee, G.D. Mahan, D. E. Eastman, R.A. Pollak, L. Ley, F.R. McFeely, S.P. Kowalczyk and D.A. Shirely, Phys. Rev. B, 11, 3614, 1975.
15. J.C.Parlebas et al. , J.Phys. France 51 , (1990) , 639-650

16. A. Szytula , B. Penc, A. Jezierski, Materials
Science-Poland, Vol.26, No. 3, 2008

Table -1 Energy separation ΔE of Rare Earth compounds

S.No.	Compounds	Z	σ	ω	Author's Calculated Plasmon Energy Separation (ΔE_s)	Experimental value of Energy Separation Ref. [15,16]
1	PrNiSb ₂	1	443.12	9.34	4.18	3.9
2	NdNiSb ₂	1	447.28	9.43	4.18	4.8
3	La ₂ CuO ₄	4	405.33	8.55	8.36	8.73
4	Nd ₂ CuO ₄	4	415.99	8.77	8.36	9.01
5	Gd ₂ CuO ₄	4	284.75	6.00	8.36	8.73
6	Pr(OH) ₃	7	191.90	4.05	11.06	11.70
7	Nd(OH) ₃	6	195.23	4.12	10.24	10.60
8	Sm(OH) ₃	4	201.35	4.24	8.36	8.50

Table -2 Relative Intensity of Rare Earth compounds

S.No.	Compounds	ΔE_s	R_s	Alpha (α)	Beta (β)	Author Calculated Relative Intensity	Experimental value of Relative Intensity Ref. [15,16]	Intensity Assignment
1	PrNiSb ₂	2.96	6.33	1.182	0.759	1.1011719	1.08	$\beta + \beta^2/2\alpha$
2	NdNiSb ₂	2.96	6.33	1.182	0.759	0.4184298	0.42	$\beta - \beta^2/2\alpha$
3	La ₂ CuO ₄	5.91	3.99	0.938	0.47864	0.37111941	0.411	$\beta - \beta^2/2\alpha$
4	Nd ₂ CuO ₄	5.91	3.99	0.938	0.47864	0.37111941	0.363	$\beta - \beta^2/2\alpha$
5	Gd ₂ CuO ₄	5.91	3.99	0.938	0.47864	0.37111941	0.388	$\beta - \beta^2/2\alpha$
6	Pr(OH) ₃	7.82	3.31	0.855	0.39719	1.73677471	1.7	$3^*(\beta + 0.1 + \beta^2/2\alpha + \beta^3/6\alpha^2)$
7	Nd(OH) ₃	7.24	3.48	0.877	0.41813	1.83197714	2	$3^*(\beta + 0.1 + \beta^2/2\alpha + \beta^3/6\alpha^2)$
8	Sm(OH) ₃	5.91	3.99	0.938	0.47864	2.12073651	2.5	$3^*(\beta + 0.1 + \beta^2/2\alpha + \beta^3/6\alpha^2)$

* Reader , Ganpat Sahai Post Graduate College , Sultanpur ,U.P. India

** Associate Professor, Rajarshi Rananjay Sinh Institute of Management & Technology, Amethi , CSJ Nagar, U.P. , India

****E-Mail ID – ajay_gspgcs@rediffmail.com**

Improvement of Power System Stability by Simultaneous AC-DC Power Transmission

T.Vijay Muni, T.Vinoditha, D.Kumar Swamy

Abstract— This paper presents the concept of simultaneous ac-dc power transmission. Long extra high voltage (EHV) ac lines cannot be loaded to their thermal limits due to this instability occurs in the power system. With the scheme proposed in this paper, it is possible to load these lines very close to their thermal limits. The conductors are allowed to carry usual ac along dc superimposed on it. The advantage of parallel ac-dc transmission for improvement of transient stability and dynamic stability and damp out oscillations have been established. Simulation study is carried out in MATLAB software package. The results shows the stability of power system when compared with only ac transmission.

Index Terms— Extra high voltage (EHV) transmission, flexible ac transmission system (FACTS), HVDC, MATLAB, simultaneous ac-dc transmission, Power System Stability, Transmission Efficiency

1 INTRODUCTION

HVDC transmission lines in parallel with EHV ac lines are recommended to improve transient and dynamic stability as well as to damp out oscillations in power system. Long EHV ac lines can not be loaded to its thermal limit to keep sufficient margin against transient instability. But for optimum use of transmission lines here is a need to load EHV ac lines close to their thermal limits by using flexible ac transmission system (FACTS) components. Very fast control of SCRs in FACTS devices like state VAR system (SVS), controlled series capacitor (CSC), static phase shifter (SPS) and controlled braking resistors oscillations as well as to control the voltage profile of the line by controlling the total reactive power flow. Only the basic idea is proposed along with the feasibility study using elementary laboratory model. The main object is to emphasize the possibility of simultaneous ac-dc transmission with its inherent advantage of power flow control improves stability and damps out oscillations in power system.

EHV ac line may be loaded to a very high value if the conductors are allowed to carry superimposed dc current along with ac current. The added dc power flow does not cause any transient instability.

- T.Vijay Muni received Masters Degree in Power & Industrial Drives from JNT University, Kakinada, India in 2010. Presently he working as Assistant Professor in Electrical and Electronics Department, Sri Sarathi Institute of Engineering & Technology, Nuzvid, India. PH-09000055144. E-mail: www.vijaymuni@gmail.com
- T.Vinoditha is currently pursuing master's degree program in Electrical Power Systems in JNT University, Hyderabad, India. PH-09052352600. E-mail: tadanki_vinoditha@yahoo.com
- D.Kumar Swamy received Masters Degree in EPSHV from JNT University, Kakinada, India. Presently he working as Associate Professor & Head in Electrical & Electronics Engineering Department, Dr. Paul Raj Engineering College, Bhadrachalem, India. PH-09866653638. E-mail: dkswamy@yahoo.co.in

This paper presents a simple scheme of simultaneous EHV ac-dc power flow through the same transmission line with an object to achieve the advantages of parallel ac-dc transmission. Simultaneous ac-dc transmission may also claim advantages in some specific applications LV (low voltage) and MV (Medium voltage) system.

The flexible ac transmission system (FACTS) concepts, based on applying state-of-the-art power electronic technology to existing ac transmission system, improve stability to achieve power transmission close to its thermal limit. Another way to achieve the same goal is simultaneous ac-dc power transmission in which the conductors are allowed to carry superimposed dc current along with ac current. Ac and dc power flow independently, and the added dc power flow does not cause any transient instability.

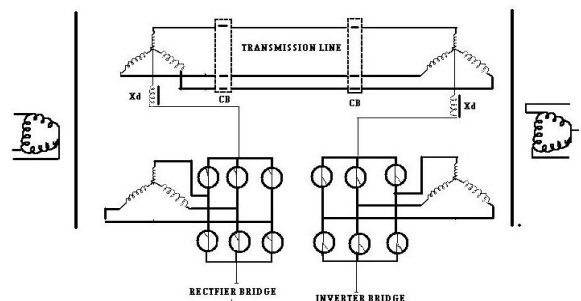


fig1 BASIC SCHEME FOR SIMULTANEOUS AC-DC TRANSMISSION

2 COCEPT OF SIMULTANEOUS AC-DC TRANSMISSION

The circuit diagram in Figure1 shows the basic scheme for simultaneous ac-dc transmission. The dc power is obtained through the rectifier bridge and in-

jected to the neutral point of the zigzag connected secondary of sending end transformer, and again it is re-converted to ac by the inverter bridge at the receiving end. The inverter bridge is again connected to the neutral of zigzag connected winding of the receiving end transformer. Star connected primary windings in place of delta-connected windings for the transformers may also be used for higher supply voltage. The single circuit transmission line carries both 3-phase ac and dc power. It is to be noted that a part of the total ac power at the sending end is converted into dc by the tertiary winding of the transformer connected to rectified bridge. The same dc power is reconverted to ac at the received end by the tertiary winding of the receiving end transformer connected to the inverter bridge. Each conductor of the line carries one third of the total dc current along with ac current I_a . The return path of the dc current is through the ground. Zigzag connected winding is used at both ends to avoid saturation of transformer due to dc current flow. A high value of reactor, X_d is used to reduce harmonics in dc current.

In the absence of zero sequence and third harmonics or its multiple harmonic voltages, under normal operating conditions, the ac current flow will be restricted between the zigzag connected windings and the three conductors of the transmission line. Even the presence of these components of voltages may only be able to produce negligible current through the ground due to high of X_d .

Assuming the usual constant current control of rectifier and constant extinction angle control of inverter, the equivalent circuit of the scheme under normal steady state operating condition is shown in Fig.2.

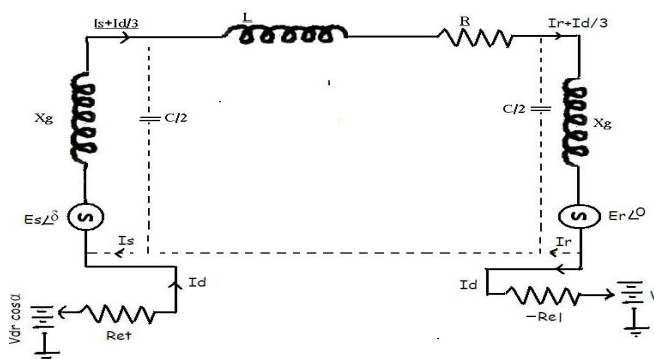


figure 2 Equivalent circuit

The dotted line in the figure shows the path of ac return current only. The ground carries the full dc current I_d only and each conductor of the line carries $I_d/3$ along with the ac current per phase

The expressions for ac voltage and current and the power equations in terms of A,B,C and D parameters of each line when the resistive drop in transformer winding and in the line conductors due to dc current

are neglected can be written as

Sending end voltage:

$$V_s = AV_R + BI_R \tag{1}$$

Sending end current:

$$I_s = CV_R + DI_R \tag{2}$$

Sending end power:

$$P_{s+JQS} = (-V_s V_R^*)/B^* + (D^*/B^*) V_s^2 \tag{3}$$

Receiving end power:

$$P_{R+JQR} = (V_s^* V_R)/B - (A^*/B^*) V_R^2 \tag{4}$$

The expressions for dc current and the dc power, when the ac resistive drop in the line and transformer are neglected,

Dc current:

$$I_d = (V_{dr} \cos \alpha - V_{di} \cos \gamma) / (R_{er} + (R/3) - R_{ci}) \tag{5}$$

Power in inverter:

$$P_{di} = V_{di} \times I_d \tag{6}$$

Power in rectifier:

$$P_{dr} = V_{dr} \times I_d \tag{7}$$

Where R is the line resistance per conductor, R_{cr} and R_{ci} commutating resistances, α and γ , firing and extinction angles of rectifier and inverter respectively and V_{dr} and V_{di} are the maximum dc voltages of rectifier and inverter side respectively. Values of V_{dr} and V_{di} are 1.35 times line to line tertiary winding ac voltages of respective sides.

Reactive powers required by the converters are:

$$Q_{di} = P_{di} \tan \theta_i \tag{8}$$

$$Q_{dr} = P_{dr} \tan \theta_r \tag{9}$$

$$\cos \theta_i = (\cos \gamma + \cos(\gamma + \mu_i)) / 2 \tag{10}$$

$$\cos \theta_r = (\cos \alpha + \cos(\alpha + \mu_r)) / 2 \tag{11}$$

Where μ_i and μ_r are commutation angles of inverter and rectifier respectively and total active and reactive powers at the two ends are

$$P_{st} = P_s + P_{dr} \text{ and } P_{rt} = P_R + P_{di} \tag{12}$$

$$Q_{st} = Q_s + Q_{dr} \text{ and } Q_{rt} = Q_R + Q_{di} \tag{13}$$

Total transmission line loss is:

$$P_L = (P_s + P_{dr}) - (P_R + P_{di}) \tag{14}$$

I_a being the rms ac current per conductor at any point of the line, the total rms current per conductor becomes:

$$I = \sqrt{I_a^2 + (I_d/3)^2} \text{ and } P_L \cong 3I^2R \tag{15}$$

If the rated conductor current corresponding to its allowable temperature rise is I_{th} and

$I_a = X * I_{th}$; X being less than unity, the dc current becomes:

$$I_d = 3 \times (\sqrt{1-X^2}) I_{th} \tag{16}$$

The total current I in any conductor is asymmetrical but two natural zero-crossings in each cycle in

current wave are obtained for $(I_d/3I_a) < 1.414$. The instantaneous value of each conductor voltage with respect to ground becomes the dc voltage V_d with a superimposed sinusoidally varying ac voltages having rms value E_{ph} and the peak value being:

$$E_{max} = V + 1.414 E_{ph}$$

Electric field produced by any conductor voltage possesses a dc component superimposed with sinusoidally varying ac component. But the instantaneous electric field polarity changes its sign twice in cycle if $(V_d/E_{ph}) < 1.414$. Therefore, higher creepage distance requirement for insulator discs used for HVDC lines are not required.

Each conductor is to be insulated for E_{max} but the line to line voltage has no dc component and $E_{LL(max)} = 2.45 E_{ph}$. Therefore, conductor to conductor separation distance is determined only by rated ac voltage of the line.

Assuming $V_d/E_{ph} = k$

$$P_{dc}/P_{ac} \cong (V_d * I_d)/(3 * E_{ph} * I_a * \cos\theta) = (k * \sqrt{1-x^2})/(x * \cos\theta) \quad (17)$$

Total power

$$P_t = P_{dc} + P_{ac} = (1 + [k * \sqrt{1-x^2}]/(x * \cos\theta)) * P_{ac} \quad (18)$$

Detailed analysis of short current ac design of protective scheme, filter and instrumentation network required for the proposed scheme is beyond the scope of present work, but preliminary qualitative analysis presented below suggests that commonly used techniques in HVDC/ac system may be adopted for this purposes.

In case of fault in the transmission system, gate signals to all the SCRs are blocked that to the bypass SCR's are released to protect rectifier and inverter bridges. CBs are then tripped at both ends to isolate the complete system. As mentioned earlier, if $(I_d/3I_a) < 1.414$, CBs connected at the two ends of transmission line interrupt current at natural current zeroes and no special dc CB is required. To ensure proper operation of transmission line CBs tripping signals to these CBs may only be given after sensing the zero crossing of current by zero crossing detectors. Else CB's connected to the delta side of transformers (not shown in figure1) may be used to isolate the fault. Saturation of transformer core, if any, due to asymmetric fault current reduces line side current but increases primary current of transformer. Delta side CBs, designed to clear transformers terminal faults and winding faults, clear these faults easily.

Proper values of ac and dc filters as used in HVDC system may be connected to the delta side and zigzag neutral respectively to filter out higher harmonics from dc and ac supplies. However, filters may be omitted for low values of V_d and I_d .

At neutral terminals of zigzag winding dc current and voltages may be measured by adopting common methods used in HVDC system. Conventional

cvts as used in EHV ac lines are used to measure ac component of transmission line voltage. Superimposed dc voltage in the transmission line does not affect the working of cvts. Linear couplers with high air-gap core may be employed for measurement of ac component of line current as dc component of line current is not able to saturate high air-gap cores.

Electric signal processing circuits may be used to generate composite line voltage and current waveforms from the signals obtained for dc and ac components of voltage and current. Those signals are used for protection and control purposes.

3 SELECTION OF TRANSMISSION VOLTAGE

The instantaneous value of each conductor voltage with respect to ground becomes more in case of simultaneous ac-dc transmission system by the amount of the dc voltage superimposed on ac and more discs are to be added in each string insulator to withstand this increased dc voltage. However, there is no change required in the conductor separation distance, as the line-to-line voltage remains unaltered. Therefore, tower structure does not need any modification if same conductor is used. Another possibility could be that the original ac voltage of the transmission be reduced as dc voltage is added such that peak voltage with respect to ground remain unchanged. Therefore, there would be no need to modify the towers and insulator strings.

4 PROPOSED APPLICATIONS

1. Long EHV ac lines can not be loaded to their thermal limit to keep sufficient margin against transient instability and to keep voltage regulation within allowable limit, the simultaneous power flow does not impose any extra burden on stability of the system, rather it improves the stability. The resistive drop due to dc current being very small in comparison to impedance drop due to ac current, there is also no appreciable change in voltage regulation due to superimposed dc current.

2. Therefore one possible application of simultaneous ac-dc transmission is to load the line close to its thermal limit by transmitting additional dc power. Figure3 shows the variation of P_t/P_{ac} for changing values of k and x at unity power factor. However, it is to be noted that additional conductor insulation is to be provided due to insertion of dc.

3. Necessity of additional dc power transmission will be experienced maximum during peak load period which is characterized with lower than rate voltage. If dc power is injected during the peak loading period only with V_d being in the range of 5% to 10% of E_{ph} , the same transmission line without having any enhanced insulation level may be allowed to be used For

a value of $x=0.7$ and $V_d = 0.05 E_{ph}$ or $0.10 E_{ph}$, 5.1% or 10.2% more power may be transmitted.

4. By adding a few more discs in insulator strings of each phase conductor with appropriate modifications in cross-arms of towers insulation level between phase to ground may be increased to a high value, which permits proportional increase in E_{max} . Therefore higher value of V_d may be used to increase dc and total power flow through the line. This modification in the exiting ac lines is justified due to high cost of a separate HVDC line.

5. With the very fast electronic control of firing angle (α) and extinction angle (γ) of the converters, the fast control of dc power may also be used to improve dynamic stability and damping out oscillations in the system similar to that of the ac-dc parallel transmission lines.

6. Control of α and γ also controls the rectifier and inverter VAR requirement and therefore, may be used to control the voltage profile of the transmission line during low load condition and works as inductive shunt compensation. It may also be considered that the capacitive VAR of the transmission line is supplying the whole or part of the inductive VAR requirement of the converter system. In pure HVDC system capacitance of transmission line cannot be utilized to compensate inductive VAR.

7. The independent and fast control of active and reactive power associated with dc, superimposed with the normal ac active and reactive power may be considered to be working as another component of FACTS.

8. Simultaneous ac-dc power transmission may find its application in some special cases of LV and MV distribution system.

When 3-phase power in addition to dc power is supplied to a location very near to a furnace or to a work place having very high ambient temperature, rectification of 3-phase supply is not possible at that location using semiconductor rectifier. In such place simultaneous ac-dc transmission is advantageous.

In air craft 3-phase loads are generally fed with higher frequency supply of about 400Hz and separate line is used for dc loads. Skin effect restricts the optimum use of distribution wires at high frequency. Simultaneous ac-dc power transmission reduces both volume and weight of distributors.

9. Another possible application is the transmission of dc power generated by PV solar cells directly to remote dc loads through 3-phase ac line. In all cases of separate dc supply filter networks are not required.

5 EXPERIMENTAL VERIFICATION

The feasibility of the basic scheme of simultaneous ac-dc transmission was verified in the laboratory. Transformer having a rating of 2 kVA, 400/230/110V

are used at each end. A supply of 3-phase, 400V, 50Hz are given at the sending end and a 3-phase, 400 V, 50 Hz, 1 HP induction motor in addition to a 3-phase, 400V, 0.7 KW resistive load was connected at the receiving end. A 10 A, 110 Vdc reactor (X_d) was used at each end with the 230V zigzag connected neutral. Two identical SCR bridges were used for rectifier and inverter. The dc voltages of rectifier and inverter bridges were adjusted between 145 V to 135 V to vary dc current between 0 to 3A.

The same experiment was repeated by replacing the rectifier at the sending end and the inverter at receiving end by 24V battery and a 5A, 25 rheostat respectively, between X_d and ground.

The power transmission with and without dc component was found to be satisfactory in all the cases. To check the saturation of zigzag connected transformer for high value of I_d , ac loads were disconnected and dc current was increased to 1.2 times the rated current for a short time with the input transformer kept energized from 400V ac. But no changes in exciting current and terminal voltage of transformer were noticed verifying no saturation even with high value of I_d .

6 SIMULATION RESULTS

The loadability of Moose (commercial name), ACSR, twinbundle conductor, 400-kV, 50-Hz, 450-km double circuit line has been computed.

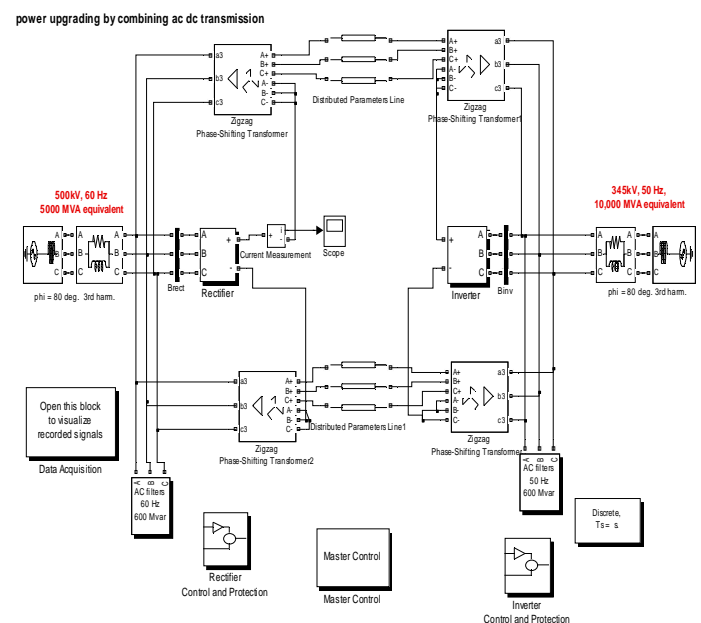


Fig 3: Simulink Model of Simultaneous AC-DC Transmission

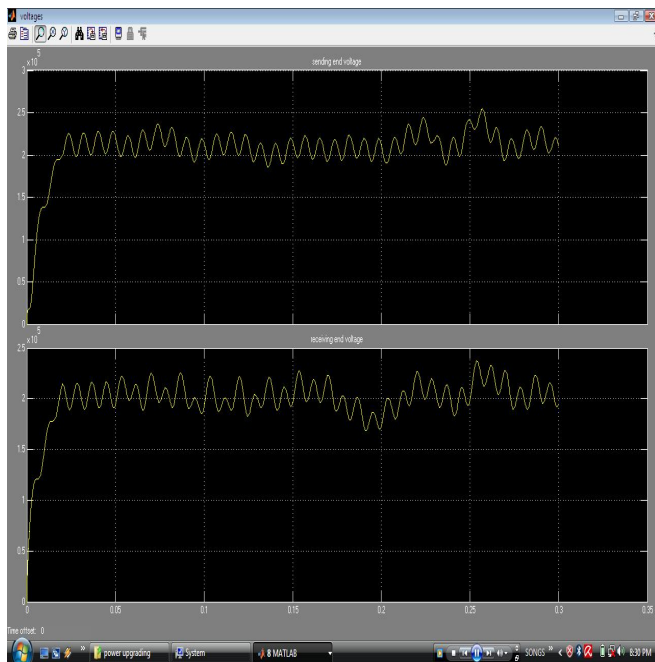


Fig 4 Sending end and receiving end voltages

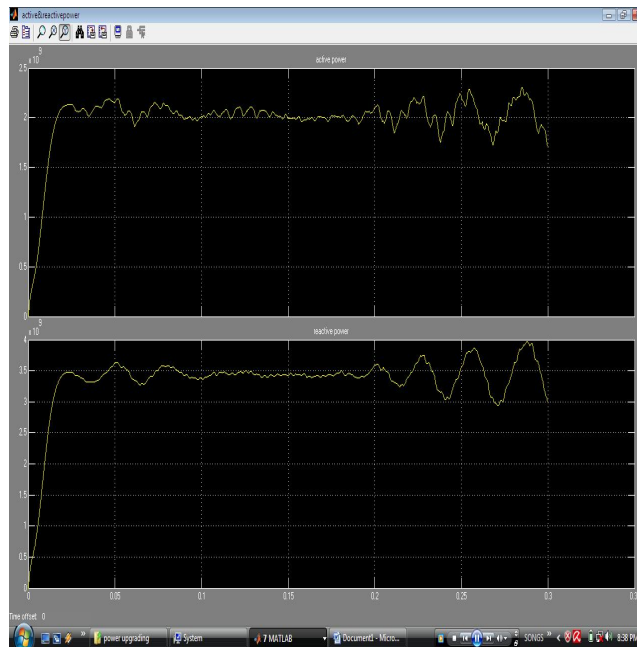


Fig 6: Combined AC-DC currents

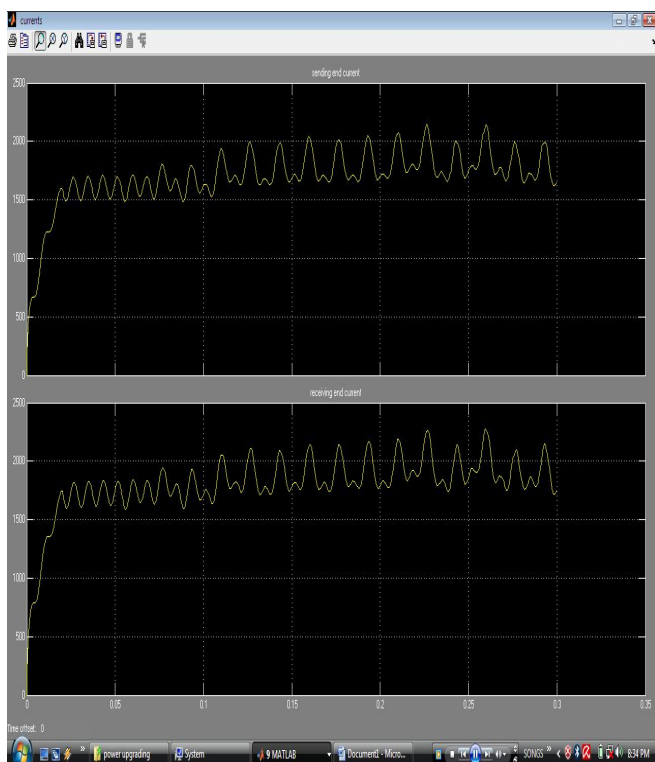


Fig 5: Sending and receiving currents

TABLE I
COMPUTED RESULTS

Power Angle	30	45	60	75
AC Current(kA)	0.416	0.612	0.80	0.98
DC Current(kA)	5.25	5.07	4.80	4.50
AC Power(MW)	290	410	502	560
DC Power(MW)	1685	1625	1545	1150
Total Power(MW)	1970	2035	2047	1710

TABLE II
SIMULATION RESULTS

Power Angle	30	45	60	75
P_s (MW)	2306	2370	2380	2342
P_{ac} (MW)	295	410	495	540
P_{dc} (MW)	1715	1657	1585	1498
P_{ac} loss (MW)	12	30	54	82
P_{dc} loss (MW)	280	265	241	217
P_R (MW)	1988	2050	2060	1995

7 CONCLUSION

A simple scheme of simultaneous EHV ac-dc power transmission through the same transmission line has been presented. Expressions of active and reactive powers associated with ac and dc, conductor voltage level and total power have been obtained for

steady state normal operating condition. The possible applications of the proposed scheme may be listed as: loading a line close to its thermal limit, improvement of transient and dynamic stability and damping of oscillations. In LV and MV distribution system the proposed scheme may be applied in a workplace having high ambient temperature or fed with high frequency supply or with PV solar cells. Only the basic scheme has been presented with qualitative assessment for its implementation. Details of practical adaptation are beyond the scope of the present work.

ACKNOWLEDGMENT

We are thankful to Department of Electrical and Electronics Engineering of Sri Sarathi Institute of Engineering and Technology, Nuzvid, India & Dr. Paul Raj Engineering College, Bhadrachalem, India with whom we had useful discussions regarding HVDC, Performance of transmissions lines. Any suggestions for further improvement of this topic are most welcome

REFERENCES

- [1] N. G. Hingorani, "FACTS—flexible A.C. transmission system," in *Proc. Inst. Elect. Eng. 5th. Int. Conf. A.C. D.C. Power Transmission*,
- [2] Padiyar. 'HVDC Power Transmission System.' Wiley Eastern, New Delhi, 1993)
- [3] H. Rahman and B H Khan " *Stability Improvement of Power Systemby Simultaneous AC-DC Power Transmission*" Electric Power System Research Journal, Elsevier, Paper Editorial ID No. EPSRD- 06-00732, Press Article No. EPSR-2560— Digital Object.
- [4] I W Kimbark. 'Direct Current Transmission Vol-I.'Wiley, New York, 1971.

A Comprehensive Distributed Generation Planning Optimization with Load Models

Mahmood Sheidaee, Mohsen Kalanter

Abstract— In this paper a comprehensive model for Distribution Systems Planning (DSP) in the case of using Distributed Generation (DG), with regard to load models is provided. Proposed model optimizes size and location of the distributed generation. This model can optimize investment cost in distributed generation better than other solutions. It minimizes the operating costs and total cost of the system losses. This Model affects the optimum location and size of the distributed generation in distribution systems significantly. Simulation studies based on a new multiobjective evolutionary algorithm is achieved. It is important that in the analysis made in this paper, DG is introduced as a key element in solving the DSP. Moreover, the proposed method easily and with little development can satisfy the other goals.

Index Terms— Economic Analysis, Distributed Generation, Distribution Systems Planning, Load Models.

1 INTRODUCTION

Distribution Companies (DISCOs) should apply new strategies in order to increase economic power generation (because of load growth) and giving services to customers and not being remained behind in market competition of electrical power. By using of new alternatives, these goals are available for solution of planning problem of distribution system in addition to traditional ones.

The load growth value is predicted in traditional options so that it may reach to a certain level. Then, a new capacity shall be added to the given system. By considering new electrical substations or via expanding the capacity of the existing substations through new feeder, such a new capacity will be obtained by both of them [1], [2]. One of the new options to planning for increase in capacity is Distributed Generation. DG may lower total costs in the system, decrease load flow within system, and improve voltage profile [3], [4], and leading to decreased system losses [3]-[5], relieving the heavy loaded feeders and increase lifetime of equipments [6].

From [7]-[10], a perfect revision is carried out on load models which are applicable to load flow and dynamic studies. Such studies are conducted on frequency or voltage dependent load models.

During recent years, the studies on evolutionary algorithms have shown that these methods have not many of previous problems [11]. In general, these methods may obtain multiple pareto optimal solutions in one single run.

This paper suggests using of DG by DISCOs, as a new economic tool for Distribution System Planning (DSP) Problem. The proposed approach makes decision on DG optimal location and size and the optimized power which should inject through distribution system. The derived results from this model may be used for bill estimation of customers of Distribution Company. Two comprehensive scenarios will be discussed to cover various probabilities.

Similarly, in this paper, the effect of load models on DG location and size planning optimization has been argued. We will see that load models considerably affect on planning for location and size of DG within distribution networks. Also, an approach is given to solve the problem which is based on Strength Pareto Evolutionary Algorithm (SPEA).

2 DISTRIBUTION SYSTEM PLANNING MODEL

In the case of load growth in power electricity market, DISCO has two options to meet such demand.

1- *Scenario A*: Purchasing the required extra power from main grid and extending the existing substations in distribution network. At this scenario, DISCO has to develop the existing substations by installation of new transformers and upgrading some existing feeders' capacities if they have not sufficient thermal capacity, and purchasing power from the main grid.

2- *Scenario B*: Investment on DG as an alternative candidate option for solving the DSP problem and purchase power from main grid and extending of the existing substation.

A. Model Formulation

This paper aims to minimize the investment and operating costs of DG, reduction of active and reactive losses, improvement in voltage profile and relieving the heavy loaded feeders. It also conducts study about impact of voltage dependent load models, namely, residential, industrial and commercial load models within different scenarios of planning. Load models are defined as follows.

$$P_i = P_{oi} |V_i|^\alpha / Q_i = Q_{oi} |V_i|^\beta \quad (1)$$

where P_i and Q_i are active and reactive power at bus i , P_{0i} and Q_{0i} are active and reactive power operating point in bus i , V_i is voltage in bus i and α and β are active and reactive power exponents. In a constant power model conventionally used in power flow studied $\alpha = \beta = 0$ is assumed. The values which are used for are active and reactive power exponents in industrial, residential and commercial load models in this paper are given in Table 1 [12].

TABLE 1
EXPONENT VALUES

Load Type	α	β
Constant	0	0
Industrial	0.18	6.00
Residential	0.92	4.04
Commercial	1.51	3.40

During studying residential model, it is assumed that the system only has residential loads. Similarly, this is also applied for industrial and commercial loads where all these loads are of industrial and commercial types, respectively. In real situations, loads aren't exactly residential, commercial and industrial, so the mixture load class should be foreseen for distribution system. There are several ideas to study on effect of DG within distribution systems. One of such ideas is the computation of multiple indices to describe the effects of disperse generations on distribution system. These indices are

1) *Active and Reactive Power Loss Indices (ILP and ILQ):*

$$ILP = \frac{P_{LDG}}{P_L} \times 100 / ILQ = \frac{Q_{LDG}}{Q_L} \times 100 \quad (2)$$

Where P_{LDG} and Q_{LDG} are total loss of active and reactive power distribution system with DG, P_L and Q_L are total loss of active and reactive power of total system without DG in the distribution network.

2) *Voltage Profile Index (IVD):* One of the advantage of proper location and size of the DG is the improvement in voltage profile.

$$IVD = \max_{i=2}^n \left(\frac{|V_1| - |V_i|}{|V_1|} \right) \times 100 \quad (3)$$

3) *MVA Capacity Index (IC):* This informational index gives information in the field of system necessities for promoting transmission line.

$$IC = \max_{i=2}^n \left(\frac{|S_{ij}|}{|CS_{ij}|} \right) \quad (4)$$

4) *Cost Index M\$ (J):*

$$J = \sum_{i=1}^M C_{fi} \left(S_{DG_i}^{\text{Max}} + BK \right) + 8760 \sum_{t=1}^T \sum_{i=1}^M \beta^t C_{ri} S_{DG_i} \quad (5)$$

$$+ 8760 \sum_{t=1}^T \beta^t \sum_{i=1}^{TN} \sum_{j=1}^M \frac{\Delta V_{ij}^2}{|Z_{ij}|} pf.C_e + A + B$$

$$A = \sum_{i=1}^{SS} \sum_{u=1}^{TU} C_{i,u} + \sum_{i=1}^{TN} \sum_{j=1}^M C_{ij} + 8760 \sum_{t=1}^T \beta^t \sum_{i=1}^{TU} pf.C_e.S_{i,u} \quad (6)$$

$$B = \sum_{i=1}^{TN} \sum_{j=1}^M C_{ij} + 8760 \sum_{t=1}^T \beta^t \sum_{i=1}^{TU} pf.C_{int}(S_{int}).S_{int}(S_{int}) \quad (7)$$

$$\beta^t = \frac{1}{(1+d)^t} \quad (8)$$

where BK denotes backup capacity of DG (MVA), β is present worth factor, C_r as investment cost in distributed generations (\$/MVA), C_r DG operating cost (\$/MVA), C_e is Electricity market price (\$/MWh), C_{ij} is Total feeder cost from i to j (\$), $C_{i,u}$ as Potential transformer u in existing substation i cost (\$), C_{int} Intertie electricity market price (\$/MWh), D is total load demand (MVA), d discount rate, i and j are bus indices, J as cost index (\$), M is total number of load buses, pf as system power factor, S_{DG} is generation power of distributed generations (MVA), S_{DG}^{Max} denotes maximum capacity of distributed generations (MVA), S_{ij} power flow in feeder connecting bus i to j (MVA), S_{ij}^{Max} is feeder thermal capacity between i and j buses (MVA), $S_{i,u}$ is dispatched power of transformer u in substation i (MVA), S_{int} as amount of power imported by the intertie (MVA), S_{SS} is power purchased by the distribution utility (MVA), S_{SS}^{Max} is capacity of existing substation (MVA), SS as number of existing substations, t incremental time intervals (year), T horizon planning year (year), TN total number of system buses, TU total number of substation transformers, V bus voltage (V), V_n system nominal voltage (V), ΔV maximum permissible voltage drop (V), and Z_{ij} feeder impedance between buses i and j (Ω ohms Ω) [13].

Optimization should be minimized by consideration of various operational constraints. Such constraints are given in (9) – (14).

1) *Total Power Conservation:* By considering losses of lines and power which is generated by DG (if it exists), sum of input and output powers should be equal to existing total load demand.

$$\sum_{i=1}^{TN} \left\{ S_{ji} - \frac{\Delta V_{ij}^2}{|Z_{ij}|} \right\} - \sum_{i=1}^M S_{ji} + S_{DG_j} = D_j, \forall j \in M \quad (9)$$

2) *Distribution Feeder's Thermal Capacity*: Distribution system's feeders have a capacity limit for the total power flow through it.

$$S_{ij} \leq S_{ij}^{\text{Max}}, \forall i \in TN, \forall j \in M \quad (10)$$

3) *Distribution Substation's Capacity*: Power which is generated by substations shall be at the substations capacity level.

$$\sum_{j=1}^M S_{SSij} \leq S_{SSi}^{\text{Max}}, \forall i \in SS, \forall j \in M \quad (11)$$

4) *Voltage Drop*: The DISCO provides the predetermined maximum permissible voltage drop limit.

$$0 \leq |V_i - V_j| \leq \Delta V, i \in TN, j \in M \quad (12)$$

5) *DG Operation*: The generated power by DG shall be lesser than DG capacity.

$$S_{DGi} \leq S_{DGi}^{\text{Max}}, \forall i \in M \quad (13)$$

6) *Intertie's Delivery Power Capacity Limit*: DISCO determines the intertie's delivery power cost rates. Rate of delivered power depends on the purchased power value by distribution network.

$$\begin{aligned} C_{\text{int}}(S_{\text{int}}) &= 1.00C_e, \forall S_{\text{int}} \in \{0,5MVA\} \\ &= 1.05C_e, \forall S_{\text{int}} \in \{5,10MVA\} \\ &= 1.15C_e, \forall S_{\text{int}} \in \{10,15MVA\} \\ &= 1.35C_e, \forall S_{\text{int}} \in \{15,20MVA\} \end{aligned} \quad (14)$$

B. Primary Distribution System under Study

The existing primary distribution system under this study is a 9-bus system [13]. This system has a 40MVA substation. Load growth which has been predicted for 4 years is approximately 28% (51.1 MVA). The given system has 4 feeders with thermal capacity of 12 MVA.

Based on market indexes in 2002 (in USD), 70 \$/MWh and 0.5 M\$/MVA have been considered as prices of electricity market and natural gas generator set, respectively. DGs have a capacity which is multiples of 1 MVA at a generated electricity price of 50 \$/MWh. An extra DG with 1 MVA capacity has been allocated for each DG as backup [13]. It is assumed that DG has unity power factor [12]. The maximum capacity of DG has been provided for each bus (maximally 4 units plus one unit as backup). This is done in order to keep distributed generation concept and DG not to be concentrated as a centralized plant so the existing main substation to be used at maximum level.

Two three phase transformers with 10 MVA power capacity and 0.2 M\$ price for each one may be also installed to expand the main substation. The cost of upgrading the existing primary distribution feeder with another of higher capacity is 0.15 M\$/Km. The price of other existing equipments and feeders will be considered zero. System power factor and discount rate are also 0.9 and 12.5%, respectively. During all optimization runs, population size and generations maximum number are 300 and 750 respectively. The maximum size of Pareto's optimal set includes 20 solutions. The probabilities of crossover and mutation are 0.9 and 0.01, respectively. Recently, the studies on evolutionary algorithms have indicated that these algorithms may be effective in removal of problems of older methods. The applied optimization algorithm is SPEA [14].

3 ANALYSIS AND RESULTS

Two argued main scenarios in this paper are simulated in order to evaluate preference of investing on DGs in solving DSP problems in comparison with other traditional planning options.

3.1 Scenario A

In this scenario, DG size is zero. This model gives the optimal cost of substation's new transformers, and the power which is dispatched by these transformers, and the amount of delivered power from main grid to distribution network for all types of load models (NL is number of lines upgrading).

Results which obtained from optimization model are given in Table 2 for all load models. It is seen that load models affect on solutions. It is possible that the obtained solution does not apply to industrial load by using of constant load model. Such an impact from load models is also observed for residential, commercial, and composed models. More money should be spent by purchase of new equipments and due to compensation for losses so these costs are added to customers' bills.

3.2 Scenario B

Table 3 indicates the solutions that are derived for investment on DG option. Comparing these results with the results of above scenario, we can see that investment on DG presents a better planning. For constant load model, four groups of DG with capacities of 4, 4, 3 and 2 MVA and 1 MVA have been selected as backup in buses 7, 9, 3 and 9 in addition a single transaction taking place by the intertie of 1.4927 MVA and the dispatched power by expanding substation is zero (we do not expand the substation). Cost of planning for this load model is 2% at this scenario, active losses 52%, and the reactive losses is 52% lesser than at first scenario. It is clear that by investing on DGs instead of purchase of power at higher

prices, DISCO may reduce its consumers' bill due to decrease in cost. Table 3 shows detailed results which are obtained for other load models in Scenario B.

DG, as a key element in DSP problem, is not used only to minimize planning cost and reduced active and reactive losses, but as it discussed, it also has several economic, social and electrical advantages. Figures (1) – (5) show voltage profiles of distribution network buses for all load models in both manners.

**TABLE 2
RESULTS OF SCENARIO A**

Index	Constant	Residential	Industrial	Commercial	Mixture
J (M\$)	31.3837	24.0529	26.7622	21.1386	25.0651
S_{int} (MVA)	4.7820	2.9595	2.2405	4.9024	3.3959
$S_{i,u}$ (MVA)	9.7132	7.6799	9.8189	4.1953	7.7625
P_L (MW)	3.06	2.55	2.74	2.36	2.63
Q_L (MVar)	2.14	1.78	1.92	1.65	1.84
NL	1	1	1	1	1

**TABLE 3
RESULTS OF SCENARIO B**

Index	Constant	Residential	Industrial	Commercial	Mixture
J (M\$)	30.7797	23.2452	26.0381	20.8633	24.7617
S_{int} (MVA)	1.4927	3.0012	0.889	4.7927	1.2087
$S_{i,u}$ (MVA)	0	0	0	0.8567	0
P_{LDG} (MW)	1.46	1.6	1.4	1.91	1.41
Q_{LDG} (MVar)	1.02	1.12	0.98	1.34	0.98
S_{DG} (MVA)	2,3,4,4	4,4	2,4,2,3	4	3,2,4,1
DG Location	5,3,9,7	9,7	3,7,5,9	9	7,3,9,5
NL	0	0	0	0	0

Voltage profiles are very better for all load models in planning with DG than without it. Load models influence on solutions.

One of the other advantages of introducing DG to solve DSP problem has been shown in Figs (6) – (10). As it seen, feeders' power flow is decreased for all load models, as a result, system losses is reduced and at last, losses cost will be decreased. Also, this reduces the feeders' load and subsequently, it increases feeders' lifetime. Similarly, we have a chance to use the existing distribution network with no need to upgrading feeders for further load growth. It is again observed that load models influence on solutions. The solutions which are derived from different load models do not apply to other models.

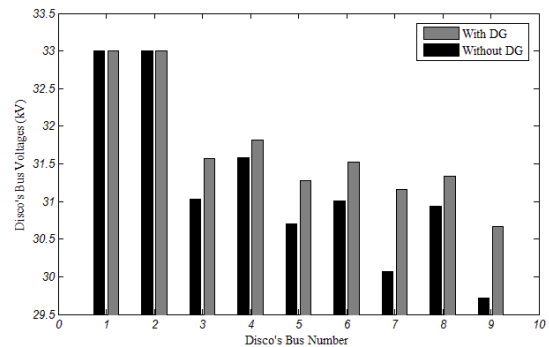


Fig. 1. DISCO's buses voltage profile for constant load model

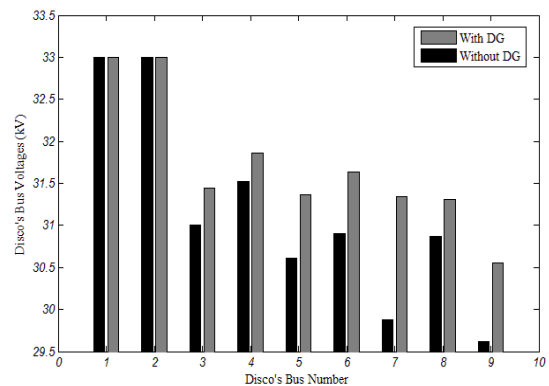


Fig. 2. DISCO's buses voltage profile for industrial load model

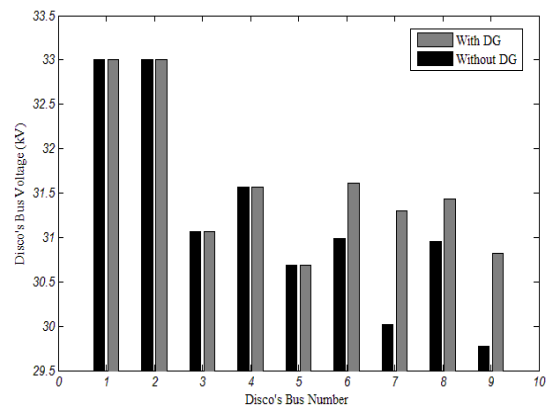


Fig. 3. DISCO's buses voltage profile for residential load model

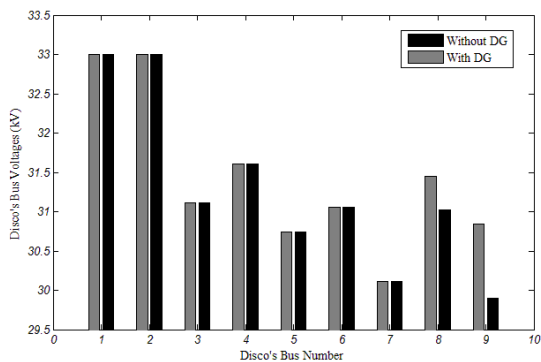


Fig. 4. DISCO's buses voltage profile for commercial load model

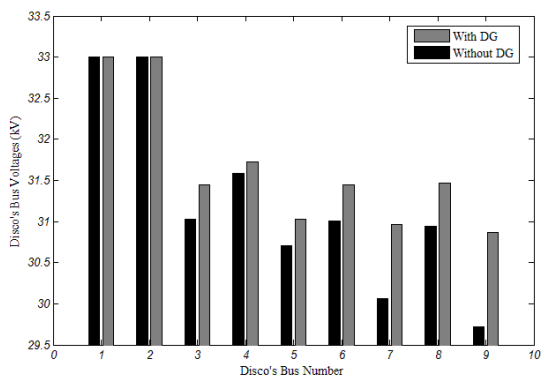


Fig. 5. DISCO's buses voltage profile for mixture load model

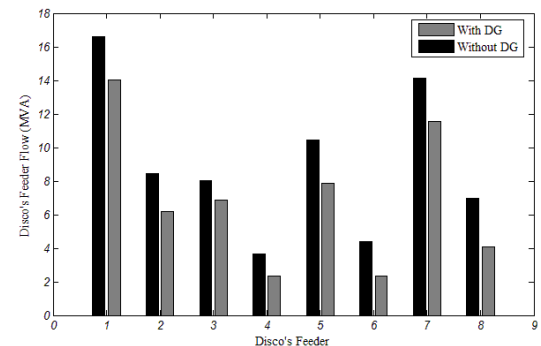


Fig. 6. DISCO's primary distribution feeder power flow for constant load model

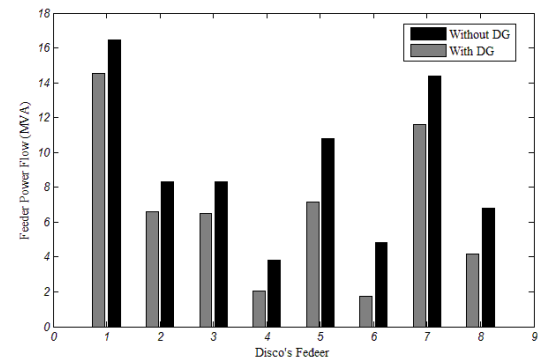


Fig. 7. DISCO's primary distribution feeder power flow for industrial load model

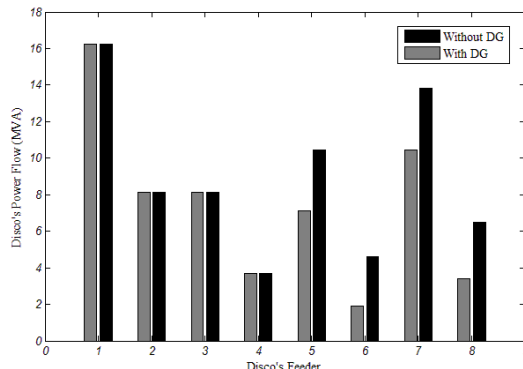


Fig. 8. DISCO's primary distribution feeder power flow for residential load model

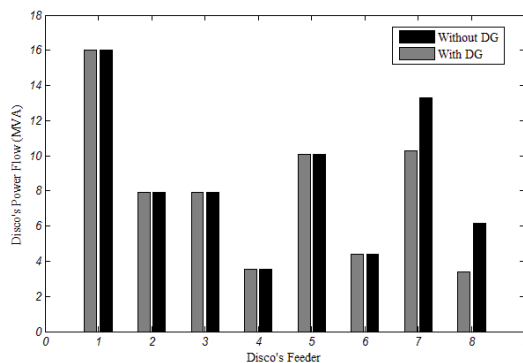


Fig. 9. DISCO's primary distribution feeder power flow for commercial load model

they are highly different from each other in cost values, losses, voltage profile and feeders' power flow, and this indicates that for an appropriate location-size planning, load models are important and crucial. As a new tool in solving DSP problem in comparison with traditional planning alternatives, investment on DGs may create further economic, social advantages. As a result, DISCO may restore its own customers and prevents these customers not to buy electricity power from other companies.

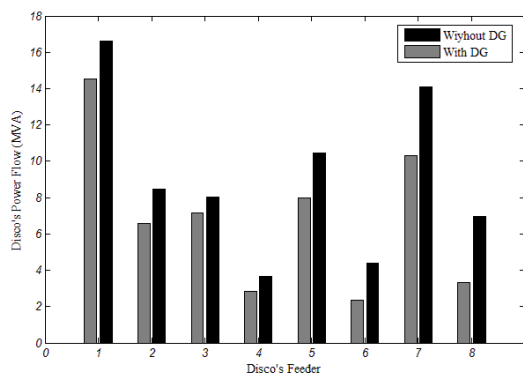


Fig. 10. DISCO's primary distribution feeder power flow for mixture load model

4 CONCLUSION

A comprehensive analysis was presented for DG location-size planning multiobjective optimization, including load models, in distribution systems. It was seen that when load models are considered, some changes occur in DG location and size. To estimate DGs optimal location and size, a new advanced optimization model was used. The proposed optimization model adapts cost, system total losses, voltage profiles and feeders' power flow and gives optimal answer which varied upon change of load models. The output results of this model not only give DG location and size, but also express a need to cost in order to purchase other new equipments (transformers and feeders upgrading). Based on information and rates that used at this paper, we observed that DG may lower planning cost, improve voltage profile of distribution system, decrease feeders' power flow of distribution networks, and minimize losses in distribution system and increase their lifetime by lower down feeders' load. Thus, DG gives an opportunity to use the existing distribution network to further load growth with no need to upgrade feeders.

REFERENCES

- [1] M. Ponnaivaikko, K. S. Prakasa, and S. S. Venkata, "Distribution system planning through a quadratic mixed-integer programming approach," *IEEE Trans. Power Del.*, vol. PWRD-2, no. 4, pp. 1157-1163, Oct. 1987.
- [2] S. K. Khator and L. C. Leung, "Power distribution planning: A review of models and issues," *IEEE Trans. Power Syst.*, vol. 12, no. 3, pp. 1151-1159, Aug. 1997.
- [3] P. P. Barker and R.W. De Mello, "Determining the impact of distributed generation on power systems. I. Radial distribution systems," in *IEEE Power Eng. Soc. Summer Meeting*, vol. 3, 2000, pp. 1645-1656.
- [4] X. Ding and A. A. Girgis, "Optimal load shedding strategy in power systems with distributed generation," in *IEEE Winter Meeting Power Eng. Soc.*, vol. 2, 2001, pp. 788-793.
- [5] N. Hadjsaid, J. F. Canard, and F. Dumas, "Dispersed generation impact on distribution networks," *IEEE Comput. Appl. Power*, vol. 12, no. 2, pp. 22-28, Apr. 1999.
- [6] L. Coles and R.W. Beck, "Distributed generation can provide an appropriate customer price response to help fix wholesale price volatility," in *IEEE Power Eng. Soc. Winter Meeting*, vol. 1, 2001, pp. 141-143.
- [7] C. Concordia and S. Ihara, "Load representation in power systems stability studies," *IEEE Trans. Power App. Syst.*, vol. PAS-101, no. 4, Apr. 1982, pp. 969-977.
- [8] IEEE Task Force on Load Representation for Dynamic Performance, "Bibliography on load models for power flow and dynamic performance simulation," *IEEE Trans. Power Syst.*, vol. 10, no. 1, Feb. 1995, pp. 523-538.
- [9] IEEE Task Force on Load Representation for Dynamic Performance, "Load representation for dynamic performance analysis," *IEEE Trans. Power Syst.*, vol. 8, no. 2, May 1993, pp. 472-482.
- [10] IEEE Task Force on Load Representation for Dynamic Performance, "Standard load models for power flow and dynamic performance simulation," *IEEE Trans. Power System*, vol. 10, no. 3, Aug. 1995, pp. 1302-1313.
- [11] C. A. C. Coello, "A comprehensive survey of evolutionary based multiobjective optimization techniques," *Knowledge and Information Systems*, vol. 1, no. 3, 1999, pp. 269-308.
- [12] D. Singh, K. S. Verma, "Multiobjective Optimization for DG Planning With Load Models," *IEEE Trans. Power Syst.*, vol. 24, no. 1, Feb. 2009, pp. 427-436.
- [13] W.El-Khattam, Y. G. Hegazy, M. M. A. Salama, "An Integrated Distributed Generation Optimization Model for Distribution System Planning," *IEEE Trans. Power System*, vol. 20, no. 2, May 2005, pp. 1158-1165.
- [14] M. A. Abido, "Environmental/Economic Power Dispatch Using Multiobjective Evolutionary Algorithms," *IEEE Trans. Power Syst.*, vol. 18, no. 4, Nov.2003, pp. 1529-1537

Positive Impact of Mobility Speed on Performance of AODV at Increased Network Load

Rajesh Deshmukh, Asha Ambhaikar

Abstract — This paper studies impact of changing mobility speed on the performance of a reactive routing protocol AODV with reference to varying network load. For experimental purposes, initially we observed the performance of AODV with increasing Network Load from 4 packets to 24 packets at the maximum mobility speed of 10 m/s. In another scenario we observed the performance of AODV with increasing Network Load from 4 packets to 24 packets at maximum mobility speed of 20 m/s. The performance of AODV is observed across Packet Delivery Ratio, Loss Packet Ratio and Routing overhead parameters. Our simulation results show that AODV is performing better with higher mobility speed at higher network load.

Index Terms— AODV, MANET, Mobility Speed, Routing, Overhead, Random Waypoint

1. INTRODUCTION

An adhoc network is a dynamic network. It allows wireless mobile nodes dynamically forming a temporary network without the use of any existing network infrastructure or centralized administration. A number of routing protocols like Dynamic Source Routing (DSR), Ad Hoc On-Demand Distance Vector Routing (AODV) and Destination-Sequenced Distance-Vector (DSDV) have been proposed. In this work an attempt has been made to compare the performance of a reactive routing protocol for mobile ad hoc networks AODV on the basis of varying number of packets with reference to mobility speed. The performance differentials are analyzed using varying mobility and packet size. These simulations are carried out using the ns-2 network simulator, which is used to run ad hoc simulations. The results presented in this paper illustrate the importance in carefully evaluating and implementing routing protocols when evaluating an ad hoc network protocol.

1. AD HOC ROUTING PROTOCOLS

Routing in Mobile Ad-hoc Network is a subject of extensive research, Because of the fact that it may be necessary to pass several hops (multi-hop) before a packet reaches the destination, a routing protocol is needed. Routing protocol has two functions, first is selection of routes for various source-destination pairs and second, Delivery of messages to their correct destination.

The second function is conceptually straightforward using a variety of protocols and data structures (routing

tables). Ad-hoc routing protocols can be classified based on different criteria. Depending upon the routing mechanism employed by a given protocol, they fall in two classes.

Table Driven Routing Protocols (Proactive): Each node in table-driven routing protocols, continuously maintains up-to-date routes to every other node in the network. Periodic routing information is transmitted throughout the network in order to maintain consistency of the routing table. Transmission occurs without delay if the route already exists, otherwise, node needs to receive routing information corresponding to its destination while traffic packets are waiting in the queue. Certain proactive routing protocols are Destination- Sequenced Distance Vector (DSDV), Wireless Routing Protocol (WRP), Global State Routing (GSR) and Cluster head Gateway Switch Routing (CGSR) [6].

On-Demand Routing Protocols (Reactive): In on demand protocols, only when a node wants to send packets to its destination it initiates a route discovery process through the network. After a route is determined or all possible permutations have been examined, the process of route discovery is completed. The discovered route has to be maintained by a route maintenance process until either the destination becomes inaccessible along every path from the source or until the route is no longer desired [6]. Some reactive protocols are Cluster Based Routing Protocol (CBRP), Ad hoc On-Demand Distance Vector (AODV), Dynamic Source Routing (DSR), Temporally Ordered Routing Algorithm (TORA), Associativity-Based Routing (ABR), Signal Stability Routing (SSR) and Location Aided

Routing (LAR) [6].

1.1 Dynamic Source Routing Protocol (DSR)

The Dynamic Source Routing (DSR) protocol is an on demand routing protocol based on source routing. DSR protocol is composed by two "on-demand" mechanisms, which are requested only when two nodes want to communicate with each other. This Protocol is composed of two essential parts of route discovery and route maintenance. Every node maintains a cache to store recently discovered paths [5]. Route Discovery and Route Maintenance are built to behave according to changes in the routes in use, adjusting them-selves when needed. Along with those mechanisms, DSR allows multiple routes to any destination, thus can lead easily to load balancing or increase robustness. In the source routing technique, a sender determines the exact sequence of nodes through which to propagate a packet. The list of intermediate nodes for routing is explicitly contained in the packet's header. In DSR [5], every mobile node in the network needs to maintain a route cache where it caches source routes that it has learned. When a host wants to send a packet to some other host, it first checks its route cache for a source route to the destination. In the case a route is found, the sender uses this route to propagate the packet. Otherwise the source node initiates the route discovery process.

1.2 Ad hoc On-Demand Distance Vector (AODV)

Ad hoc on demand distance vector (AODV) routing protocol creates routes on-demand. In AODV, a route is created only when requested by a network connection and information regarding this route is stored only in the routing tables of those nodes that are present in the path of the route [1]. AODV is a reactive protocol based upon the distance vector algorithm. The algorithm uses different types of messages to discover and maintain links. Whenever a node wants to try and find a route to another node it broadcasts a Route Request (RREQ) to all its neighbors [2]. In this protocol, each terminal does not need to keep a view of the whole network or a route to every other terminal. Nor does it need to periodically exchange route information with the neighbor terminals. Furthermore, only when a mobile terminal has packets to send to a destination does it need to discover and maintain a route to that destination terminal. In AODV, each terminal contains a route table for a destination [5]. A route table stores the following information: destination address and its sequence number, active neighbors for the route, hop count to the destination, and expiration time for the table. The expiration time is updated each time the route is used. If this route has not been used for a specified period of time, it is discarded [7].

1.3 Destination Sequenced Distance-Vector

Routing (DSDV)

The Destination Sequenced Distance Vector Protocol (DSDV) is a proactive, distance vector protocol which uses the Bellman - Ford algorithm [4]. DSDV is a hop by hop distance vector routing protocol, wherein each node maintains a routing table listing the "next hop" and "number of hops" for each reachable destination. This protocol requires each mobile station to advertise, to each of its current neighbors, its own routing table (for instance, by broadcasting its entries). The entries in this list may change fairly dynamically over time, so the advertisement must be made often enough to ensure that every other mobile computer can almost always locate every other mobile computer of the collection. In addition, each mobile computer agrees to relay data packets to other computers upon request. This agreement places a premium on the ability to determine the shortest number of hops for a route to a destination we would like to avoid unnecessarily disturbing mobile hosts if they are in sleep mode. In this way a mobile computer may exchange data with any other mobile computer in the group even if the target of the data is not within range for direct communication. DSDV requires a regular update of its routing tables, which uses up battery power and a small amount of bandwidth even when the network is idle [4].

2. MOBILITY MODEL

2.1 Random Waypoint Mobility Model

The Random waypoint model is a random-based mobility model used in mobility management schemes for mobile communication systems. Random Waypoint (RW) model assumes that each host is initially placed at a random position within the simulation area [3]. The mobility model is designed to describe the movement pattern of mobile users, and how their location, velocity and acceleration change over time [3]. Mobility models are used for simulation purposes when new network protocols are evaluated. In random based mobility simulation models, the mobile nodes move randomly and freely without restrictions. To be more specific, the destination, speed and direction are all chosen randomly and independently of other nodes. This kind of model has been used in many simulation studies. Two variants, the Random walk model and the Random direction model are variants of the Random waypoint model.

In this model, a mobile node moves from its current location to a randomly chosen new location within the simulation area, using a random speed uniformly distributed between $[v_{min}, v_{max}]$ [3]. v_{min} refers to the minimum speed of the simulation, v_{max} to the maximum speed [3]. The Random Waypoint Mobility

Model includes pause times when a new direction and speed is selected. As soon as a mobile node arrives at the new destination, it pauses for a selected time period (pause time) before starting traveling again. A Mobile node begins by staying in one location for a certain period of time (i.e. pause). Once this time expires, the mobile node chooses a random destination in the simulation area and a speed that is uniformly distributed between $[v_{min}, v_{max}]$. The mobile node then travels toward the newly chosen destination at the selected speed. Upon arrival, the mobile node pauses for a specified period of time starting the process again. The random waypoint model is the most commonly used mobility model in the simulation of ad hoc networks. It is known that the spatial distribution of network nodes moving according to this model is non-uniform. However, a closed-form expression of this distribution and an in depth investigation is still missing. This fact impairs the accuracy of the current simulation methodology of ad hoc networks and makes it impossible to relate simulation based performance results to corresponding analytical results. To overcome these problems, it is presented a detailed analytical study of the spatial node distribution generated by random waypoint mobility. It is considered that a generalization of the model in which the pause time of the mobile nodes is chosen arbitrarily in each waypoint and a fraction of nodes may remain static for the entire simulation time [3].

3. THE TRAFFIC AND SCENARIO GENERATOR

Continuous bit rate (CBR) traffic sources are used. The source-destination pairs are spread randomly over the network. The simulation uses Random Waypoint mobility model in a 1020 m x 1020 m field with varying network load of 4 packets to 24 packets whereas mobility speed is kept at 10 m/s maximum. In the next simulation network load is varied from 4 packets to 24 packets, but this time mobility speed is kept 20 m/s maximum. Here, each packet starts its journey from a random location to a random destination with a randomly chosen speed. Once the destination is reached, another random destination is targeted after a pause. The pause time, which affects the relative speeds of the mobile hosts, is kept at 20s. Simulations are run for 100 simulated seconds.

4. PERFORMANCE METRICS

Following important metrics are evaluated-

1. Packet Delivery ratio (PDR) - Packet delivery ratio is calculated by dividing the number of packets

received by the destination through the number of packets originated by the CBR source.

2. Loss Packet Ratio (LPR) - Loss Packet Ratio is calculated by dividing the number of packets that never reached the destination through the number of packets originated by the CBR source.
3. Routing Overhead - Routing overhead, which measures the ratio of total routing packets sent and the total number of packets sent.

5. SIMULATION SETUP

In this simulation we wanted to investigate how mobility speed affects on the behavior AODV with increasing network load.

TABLE 1
EVALUATION WITH MOBILITY SPEED 10 M/S

Parameter	Value
Protocols	AODV
Simulation Time	100 s
Number of Nodes	100
Network Load	4, 8, 12, 16, 20, 24 Packets
Pause Time	20 s
Environment Size	1020 m x 1020 m
Traffic Type	Constant Bit Rate
Maximum Speed	10 m / s
Mobility Model	Random Waypoint
Network Simulator	NS 2.33

TABLE 2
EVALUATION WITH MOBILITY SPEED 20 M/S

Parameter	Value
Protocols	AODV
Simulation Time	100 s
Number of Nodes	100
Network Load	4, 8, 12, 16, 20, 24 Packets
Pause Time	20 s
Environment Size	1020 m x 1020 m
Traffic Type	Constant Bit Rate
Maximum Speed	20 m / s
Mobility Model	Random Waypoint
Network Simulator	NS 2.33

6. RESULTS AND DISCUSSIONS

During the simulation we have increased the network load with maximum mobility maximum speed of 10 m/s and recorded the performance of AODV. We did this simulation for 100 simulated seconds with maximum 8 cbr connections. Readings were taken for different network loads (4, 8, 12, 16, 20 and 24 packets). Again same simulation is performed, but this time with maximum speed of 20 m/s. From the results it is evident that AODV starts to perform better with mobility speed of 20 m/s as compared to 10 m/s for same scenario. At higher network load and maximum speed of 20 m/s, the Packet Delivery ratio increases,

Loss Packet Ratio decreases and Routing Overhead decreases.

7. PERFORMANCE EVALUATION

Observation for Mobility Speed of 10 m/s: Simulation result in figure 1 shows that performance of AODV in terms of Packet Delivery Ratio degrades as network load is increased. When network load reach 12 packets, PDR is dropped considerably. Even though PDR starts to improve gradually from that point and reach a much better performance around 16 packets of load. Once again performance starts degrading, and continues to degrade more. Same with the Routing overhead, Figure 3 shows that Routing overhead keeps on increasing until network load of 12 packets and from that point overhead starts to decrease till the network load reaches 16 packets. After this point routing overhead keeps on increasing and never recovers again.

Observation for Mobility Speed of 20 m/s: Simulation result in figure 1 shows that performance of AODV degrades as network load is increased. A point to notice is that when network load reach 12 packets, performance of AODV is much improved as compared to performance with Mobility Speed of 10 m/s. Packet Delivery Ratio stays consistent until network load reaches 16 packets, even though it is performing poor than the earlier simulation scenario. PDR keeps on decreasing until a point where network load reach 20 packets. From this point PDR starts to improve gradually and achieves a much better performance as compared to performance with mobility speed of 10 m/s. About Routing overhead, Figure 3 shows that Routing Overhead remains either equal or better than 10 m/s scenario until network load reach 12 packets. Routing overhead stays consistent till 16 packets and then again gets worse till the 16 packets mark. From their AODV starts to improve the performance and achieves better readings compared to reading with 10 m/s as network load crosses the 20 packets mark.

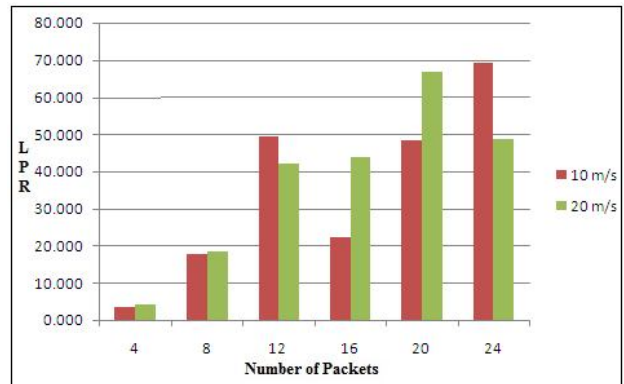


Fig 2. Number of Packets Vs Loss Packet Ratio

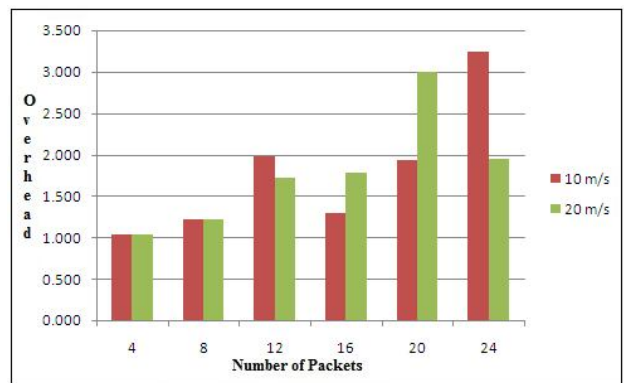


Fig 3. Number of Packets Vs Routing Overhead

8. CONCLUSION AND FUTURE WORK

Empirical results illustrate that the performance of AODV varies widely across different network loads, and study results from two different scenarios shows that increasing the mobility speed does help to improve the performance of AODV when it comes to higher network loads. Hence we have to consider the network load of an application while selecting the mobility speed.

The future scope is to find out what factors can bring more improvements in performance of AODV not only while the network load is further increased but also on the load where AODV has not performed well in simulations presented here. Further simulation needs to be carried out for the performance evaluation with not only increased mobility speed but also varying other related parameters like Pause Time, Mobility models etc.

REFERENCES

[1] Humaira Nishat, Vamsi Krishna K, Dr. D.Srinivasa Rao, Shakeel Ahmed, "Performance Evaluation of On Demand Routing Protocols AODV and Modified AODV (R-AODV) in MANET", International Journal of Distributed and Parallel Systems (IJ DPS)

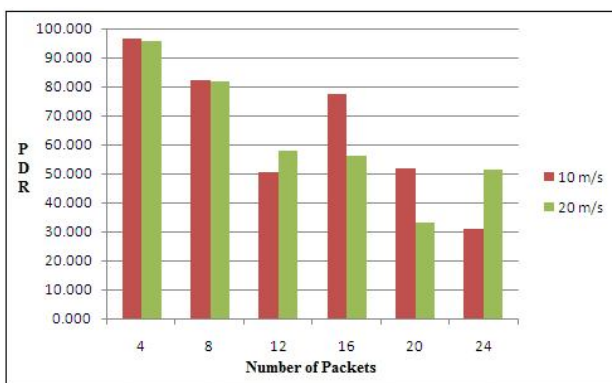


Fig 1. Number of Packets Vs Packet Delivery Ratio

- Vol.2, No.1, January 2011,
<http://airccse.org/journal/ijcdps/papers/0111ijcdps08.pdf>
- [2] Harminder S. Bindra, Sunil K. Maakar, A.L.Sangal, "Performance Evaluation of Two Reactive Routing Protocols of MANET using Group Mobility Model", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, No 10, May 2010, <http://www.ijcsi.org/papers/7-3-10-38-43.pdf>
- [3] Valentina Timcenko, Mirjana Stojanovic, Slavica Bostjancic Rakas, "MANET Routing Protocols vs. Mobility Models: Performance Analysis and Comparison", Proceedings of the 9th WSEAS International Conference on Applied Informatics and Communications (AIC '09), <http://www.wseas.us/e-library/conferences/2009/moscow/AIC/AIC47.pdf>
- [4] Abdul Hadi Abd Rahman, Zuriati Ahmad Zukamain, "Performance Comparison of AODV, DSDV and I-DSDV Routing Protocols in Mobile Ad Hoc Networks", European Journal of Scientific Research, ISSN 1450-216X Vol.31 No.4 (2009), pp.566-576, http://www.eurojournals.com/ejsr_31_4_07.pdf
- [5] Sunil Taneja, Ashwani Kush, Amandeep Makkar, "Experimental Analysis of DSR, AODV using Speed and Pause time", International Journal of Innovation, Management and Technology, Vol. 1, No. 5, December 2010, ISSN: 2010-0248, <http://www.ijimt.org/papers/79-M466.pdf>
- [6] Anuj K. Gupta, Dr. Harsh Sadawarti, Dr. Anil K. Verma, "Performance analysis of AODV, DSR & TORA Routing Protocol", IACSIT International Journal of Engineering and Technology, Vol.2, No.2, April 2010, ISSN: 1793-8236, <http://www.ijetch.org/papers/125--T203.pdf>
- [7] Parma Nand, S.C. Sharma, Rani Astya, "Simulation Based Parametric Analysis of AODV Protocol for Adhoc Network", International Journal of Advanced Engineering & Applications, Jan. 2010, <http://steps-india.com/ijaea/2.pdf>

Engineering from Chhattisgarh Swami Vivekanand Technical University, Bilai. Currently he is working as Assistant Professor in Shri Shankaracharya Institute of Professional Management & Technology (Department of Computer Science Engineering), Raipur, India since 2010. At present he is carrying out his research work in Mobile Ad hoc Network Routing Protocols.

Asha Ambhaikar received B.E from Nagpur University, Nagpur, India, in Electronics Engineering in the year 2000 and later did her M.Tech in Information Technology Allahabad Deemed University, India. Currently she is working as an Assistant Professor in Rungta College of Engineering & Technology (Department of Information Technology), Bilai, India. She has published more than 14 research papers in reputed national and international journals & conferences. Her area of interest includes, Computer Networking, Data warehousing and mining, Signal processing, Image processing, and Information systems and Security.

ABOUT AUTHORS



Rajesh Deshmukh received B.Tech in Information Technology Engineering from Allahabad Deemed University in the year 2007, and is final semester student of M.Tech in Computer Science



Development and Physical, Chemical and Mechanical Characterization of Doped Hydroxyapatite

Promita Bhattacharjee, Howa Begam, Abhijit Chanda

Abstract— In this study, we made an attempt to synthesize doped bioactive hydroxyapatite (HAp) ceramic powder using a simple Chemical method and studied its physical and mechanical properties. Different quantities (2wt% and 5 wt%) of Magnesium chloride Hexahydrate, Zinc oxide, Titanium oxide were incorporated as dopants into Hap at the time of synthesis. The synthesized powder samples were analyzed for their phases using X-ray diffraction technique, Fourier Transform Infrared Spectroscopy. The synthesized powders were uniaxially compacted and then sintered at 1250°C for 1hr in air. Vicker's hardness testing was performed to determine the hardness of the sintered structures. Fracture toughness of sintered samples was calculated using Inverted Optical Microscope with Image Analysis software.

Index Terms— Dopants, Fracture toughness, FTIR, Hydroxyapatite, Sintered structure, Vicker's hardness, XRD.

1 INTRODUCTION

Among different forms of calcium phosphates, the bioactive hydroxyapatite ($\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$) phase has been most extensively researched due to its outstanding biological responses to the physiological environment. Hydroxyapatite is brittle in nature and load bearing capacity and strength of HAp is low so we could not use it in load bearing implant (total bone replacement) where tensile stress is developed. To overcome the above stated limitations of Hap, many researchers were tried to generate nano grained HAp powder [1, 2] in different methods. There is a significant difference of properties between natural and apatite crystals found in bone mineral and the conventional synthetic HAp. Bone crystals are formed in a biological environment through the process of biomineralization. In addition, the bone mineral also contains trace ions like Na^+ , Mg^{++} , K^+ , which are known to play a important role in overall performance [1]. It has also been shown that the bioactivity of conventional synthetic HAp ceramics is inferior to the bone [1, 3-8]. During recent years, many researchers were tried in developing HAp powder doped with metallic ions by Ball milling, dry and wet milling process to increase the strength and ductility of HAp powders [1,3]. In this study, we have used a simple chemical route process that could produce HAp powder with a fairly short synthesis time. We have introduced three metal ions, in different weight percentage, which are known to be present as the bone

mineral during synthesis of HAp powder. This paper presents the synthesis and characterization of physical, mechanical and crystal structure of pure and doped HAp ceramic in detail.

2 EXPERIMENTAL PROCEDURE

2.1 Materials and methods

Pure and doped HAp powders were synthesized through water based Chemical route method. In this method, Calcium hydroxide [$\text{Ca}(\text{OH})_2$](MERCK,INDIA) and Orthophosphoric Acid [$\text{H}_3(\text{PO}_4)$] (MERCK,INDIA) were used as raw materials to produce apatite particles .



The apatite powder produced was aged for 24 hr. Then, the apatite particles in the suspension were filtered, washed with ethanol three times, and dried at 100°C for 24 hr in air. The dried powder was ground with a mortar and pestle into fine powder and subjected to calcinations at 800°C temperature for 2 hrs using electrically heated furnace (NASKAR & Co., Model No.-EN170QT) at a constant heat rate of 5°C/min, followed by cooling inside the furnace. In order to synthesize HAp powder doped with Magnesium, Zinc and Titanium, measured quantities of Magnesium Chloride Hexahydrate ($\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$, MERCK, INDIA, 96% pure), Zinc Oxide (ZnO , MERCK, INDIA, 99% pure) and Titanium Oxide (TiO_2 , MERCK, INDIA, 98% pure) were incorporated into the Calcium Hydroxide suspension before the addition of $\text{H}_3(\text{PO}_4)$ solution separately. The dopants were used in the amount of 2wt% and 5 wt% to see their effects on powder morphology and properties of the sintered ceramics.

- Promita Bhattacharjee, pursuing masters degree program in Biomedical Engineering in Jadavpur University, India.
E-mail: promita.ju@gmail.com
- Dr. Abhijit Chanda, joint director of School of Bioscience and Engineering Department of Jadavpur University, India,
E-mail: abhijitchanda.biomed@gmail.com

The resultant powders were then calcined at 800°C for 2 hr. in a electrically heated furnace. A constant heating rate of 5°C/min. was used, followed by cooling inside the furnace. Various composition of doped HAp is presented in Table I.

TABLE I

ABBREVIATION, AMOUNT (WT%), COMPOSITION OF ADDITIVES AND PERCENTAGE OF ADDITION OF METALLIC IONS

Abbreviation	Amount of additives (wt%)	Composition of additives	% of addition of metallic ion
Pure HAp	—	—	—
A 2.0	2%	MgCl 6H ₂ O	0.239%
A 5.0	5%	MgCl 6H ₂ O	0.599%
B 2.0	2%	ZnO	1.6%
B 5.0	5%	ZnO	4.02%
C 2.0	2%	TiO ₂	1.198%
C 5.0	5%	TiO ₂	2.997%

3. POWDER CHARACTERIZATION

3.1 X-ray diffraction

X-ray diffraction (XRD) technique was used to study the effect of calcination temperature and dopants on the phase evolution and phase identification. The dried calcined powders at 800°C were ground into fine powder using a mortar and pestle to break down the powder agglomerates before analyzing in an X-ray diffractometer. Powder samples were placed in the specimen holder of a Rigaku diffractometer (Model-Miniflex, Rigaku Co., Tokyo, Japan) separately, and then analyzed, using K β filtered Cu K α radiation in the step scanning mode with tube voltage of 30KV and tube current of 15mA. The XRD patterns were recorded in the 2 θ range of 0-80° with scan speed 1deg/min. The calcined powders doped with 2 wt% and 5 wt% of [MgCl₂, 6H₂O] ZnO and TiO₂, separately, were also analyzed for their phases in the same manner.

3.2. Fourier transform infrared spectroscopy

Fourier Transform Infrared Spectroscopy (FTIR) relies on the fact that most molecules absorb light in the infrared region of the electromagnetic spectrum, this absorption corresponds specifically to the bonds present in the molecule. In our experiment, we have done FTIR of as prepared calcined HAp powder and calcined HAp powder doped with different wt% of [MgCl₂, 6H₂O] ZnO and TiO₂. FTIR measurement were performed in mid IR region (5000-400 cm⁻¹) using KBr pallets in a Perkin- Elmer, Model No- 1615 (USA) instrument.

3.3. Powder Compaction And Sintering Study

As synthesized pure and doped crystalline HAp powders were uniaxially pressed using a steel mold having an internal diameter of 12-13 mm at a pressure of 173MPa, with a 2-ton press for 2 min from PEECO hydraulic pressing machine (PEECO Pvt Ltd, M/C NO.-3/PR-2/HP-1/07-08). Green ceramic structures were measured for their density and then sintered in a chamber furnace at 1250°C for 1 hr at a constant heating rate of 5°C/min.

A sintering cycle was developed to achieve better densification and to avoid cracking in the sintered specimens, by introducing several soaking temperature and tailoring the rate of heating and cooling.

Sintered ceramic structures were measured for their density and then subjected to mechanical characterization. Sintered samples intended for densification study, hardness measurement. Green and sintered ceramic specimens were measured for their geometric density from the ratio between the mass of specimen and its volume (determined by dimensional measurement).

3.4. Mechanical Characterization

Hardness Test was carried out using a Vickers diamond indenter on an automated hardness tester (Model No-LV-700AT, LECO Co, MI). During the hardness test, a load of 0.3 Kgf, 1Kgf and 3Kgf was applied on two samples of each of the composition type were tested for their hardness at three different locations with three different loads. The average of these readings were computed, reported and compared.

To determine the Fracture Toughness, we used Inverted Optical Microscope (OLYMPUS Co. Ltd., Model No-GX51F) interfacing with computer. We measured the average crack length (c) using Image Analysis Software. The cracks were developed by the indentation at the time of hardness testing. Fracture Toughness (K_{1c}) was calculated using simple equation considering radial-median crack geometry:-

$$K_{1c} = 0.016(E/H)^{1/2} P/(c)^{3/2}$$

Two samples of each of the composition type were measured the fracture toughness.

4. RESULT AND DISCUSSION

4.1. Phase Identification

4.1.1 XRD Analysis

X-ray diffraction data of the amorphous powder, powder calcined at 800°C temp. and Mg, Zn and Ti doped HAp powder at different wt% calcined at 800 °C, were recorded in 2θ range. The obtained XRD patterns of pure HAp and HAp powder doped with different wt% of Mg, Zn and Ti were presented in Fig1. The pure HAp powder calcined at 800 °C exhibited several high intensity peaks corresponding to various planes of HAp i.e. (0 0 2), (2 1 1), (1 1 2), (3 0 0) and (2 0 2) as revealed by our analysis, with reference PDF card No.- 74-0566 for hydroxyapatite. Compared to XRD pattern of pure HAp, c/a ratio of HAp crystal structure and % of volumetric strain of HAp crystals (Table: I) changed due to addition of Mg, Zn and Ti. No phase change was noticed in their XRD patterns. XRD figures 1(a),(b),(c),(d) shows the XRD traces of the dried amorphous and calcined powders, with and without dopants. The XRD patterns of the pure HAp and doped powders calcined at 800°C clearly show the presence of the most prominent peak at 2θ angle of ~31.7°C, corresponding to hydroxyapatite (2 1 1) plane. Presence of this broad peak suggests that crystallites of HAp phase were formed as a result of calcinations at 800°C. Almost identical patterns were recorded for all these compositions, which suggest that the presence of 2 wt% and 5 wt% of [MgCl₂·6H₂O] ZnO and TiO₂ as dopants did not alter the phase purity of crystalline HAp.

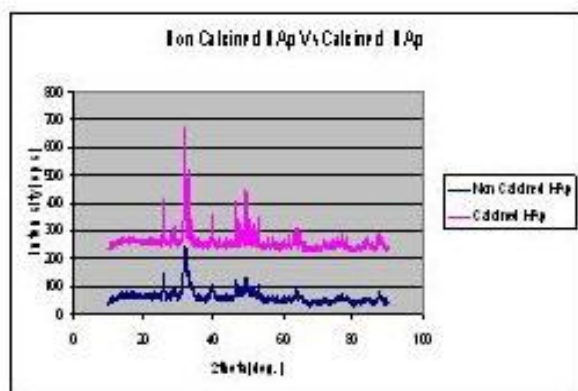


Figure 1(a). XRD patterns of Non Calcined HAp and Non Calcined HAp

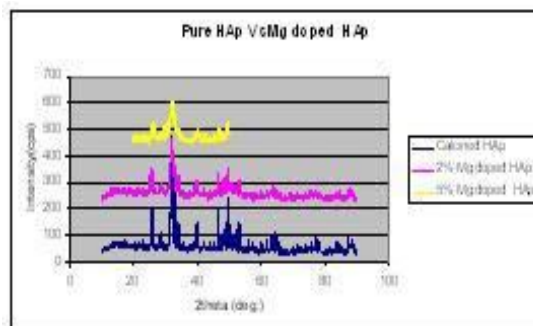


Figure 1(b). XRD patterns of Calcined HAp and Mg doped HAp(2wt% and 5wt%)

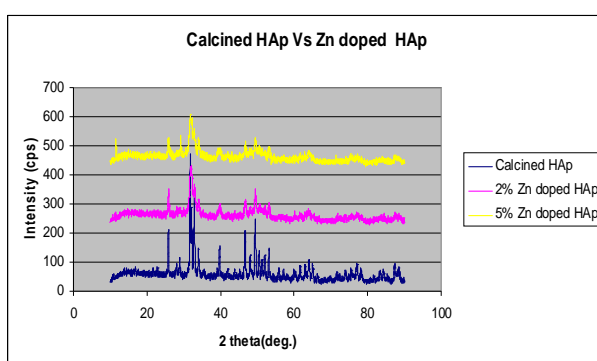


Figure 1(c). XRD patterns of Calcined HAp and Zn doped HAp(2wt% and 5wt%)

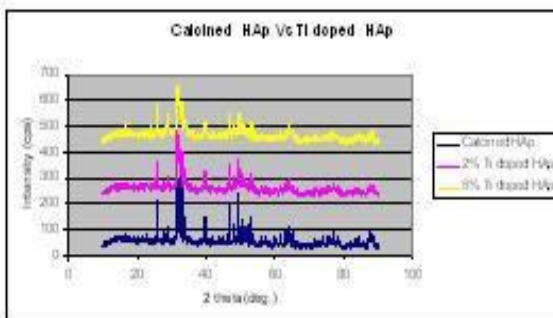


Figure 1(d). XRD patterns of Calcined HAp and Ti doped HAp(2wt% and 5wt%)

Atomic radius of Mg (150 picometer), Zn (135 picometer) and Ti (144.8 picometer) are different so they changed c/a ratio by replacing cations and developed volumetric strain.

TABLE II

c/a, UNIT CELL VOLUME AND PERCENTAGE OF VOLUMETRIC STRAIN

Composition	c/a ratio of a crystal	Unit cell volume (\AA^3)	% Volumetric strain
Pure HAp (before Calcination)	0.731	529.672	—
Pure HAp (after Calcination)	0.731	531.697	—
5% Mg doped HAp (calcined)	0.734	533.853	0.789% increase
2% Mg doped HAp (calcined)	0.729	527.725	0.368% decrease
5% Zn doped HAp (calcined)	0.739	535.059	1.017% increase
2% Zn doped HAp (calcined)	0.733	531.0451	0.259% increase
5% Ti doped HAp (calcined)	0.720	526.323	0.632% decrease
2% Ti doped HAp (calcined)	0.720	526.323	0.632% decrease

4.1.2. FTIR ANALYSIS:

The obtained FTIR patterns of pure HAp and HAp powder doped with different wt% of Mg, Zn and Ti were presented in Fig 2. The Fig 2(a) shows that FTIR of pure HAp powder calcined at 800°C. In this FTIR plot of pure calcined HAp powder we got bulge shaped peak at 3436 which reveals the presence of hydroxyl group. The Fig 2(b) shows that FTIR of 5wt% and 2wt% Mg doped HAp powder calcined at 800°C. In these two FTIR plots hydroxyl group peak is very short and tending to disappear. This may be due to the evaporation of the surface moisture of the powder. The Fig 2(c) shows that FTIR of 5wt% and 2wt% Zn doped HAp powder calcined at 800°C. In these two FTIR plots we got bulge shaped peak at 3436 which shows the presence of hydroxyl group. FTIR spectra of HAp powder and ZnO doped HAp powder present almost similar bands. The Fig 2(d) shows that FTIR of 5wt% and 2wt% Ti doped HAp powder calcined at 800°C.

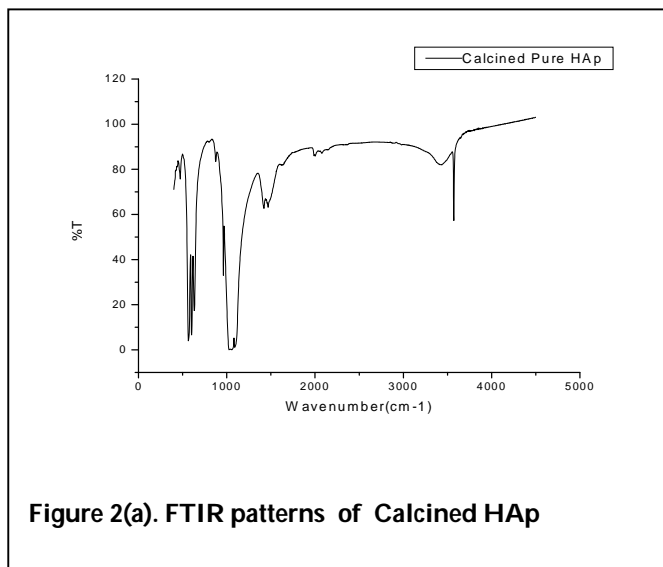


Figure 2(a). FTIR patterns of Calcined HAp

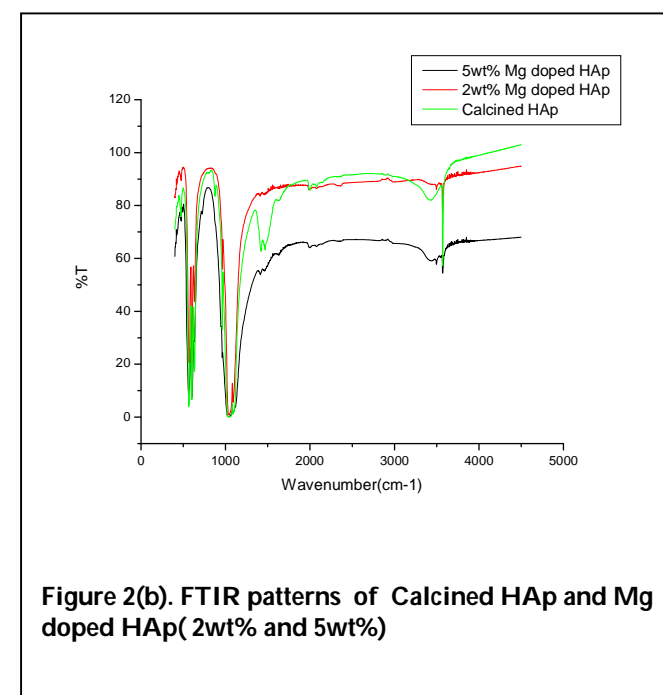
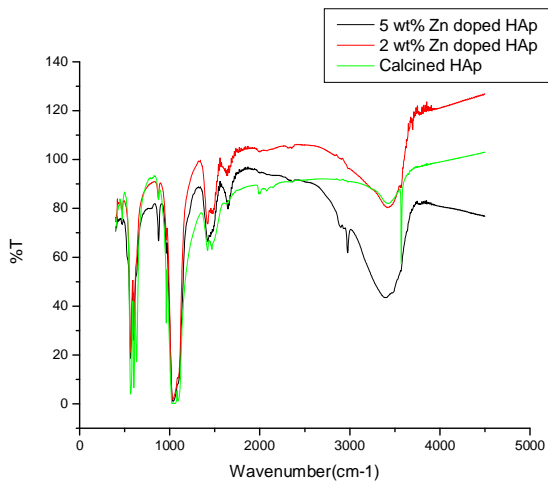


Figure 2(b). FTIR patterns of Calcined HAp and Mg doped HAp (2wt% and 5wt%)



2.(c)

Figure 2(c). FTIR patterns of Calcined HAp and Zn doped HAp(2wt% and 5wt%)

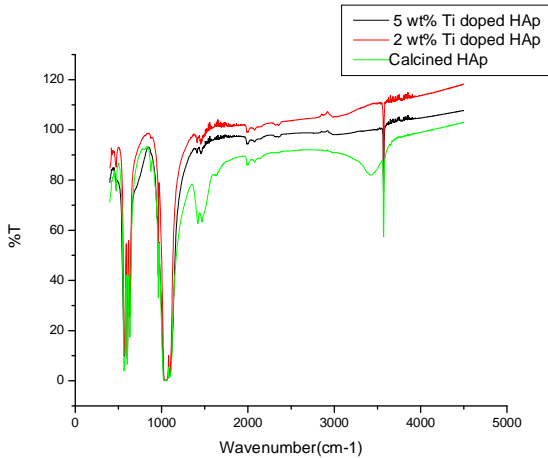


Figure 2(d). FTIR patterns of Calcined HAp and Ti doped HAp(2wt% and 5wt%)

4.2.Sintering study

Green ceramic structure prepared via uniaxial pressing were measured for their green density and were subjected to pressureless sintering. Average green density of all compositions are presented in Table III .The average diameters of all specimens before sintering is 12.15 ± 0.05 mm and thickness 5.4 ± 0.05 mm. Each sintered specimen was measured for its density and linear ,diametric and volumetric shrinkage

and average of each of these compositions were calculated and presented in Table IV. Average diameter of sintered specimens is 10 ± 0.05 mm and thickness 4.5 ± 0.05 mm. There was no major variation in shrinkage among the various compositions. Results of our sintering study showed that sintering at 1250°C helps in the densification of pure and doped crystalline HAp ceramics which is in line with literature on conventional HAp sintering [5,6,10].We also performed some sintering study at 1200°C which showed relatively lower densification(85. 25%).Sintering studies were not done above 1300°C , as it is already established that sintering of HAp ceramics above 1300°C leads to significant phase change, unwanted grain growth and deterioration of their properties. An average sintered density of 2.9 g/cc were recorded in pure HAp specimens sintered at 1250°C which is equivalent to 93.9% of theoretical density of HAp. In this work, presence of Mg, Zn and Ti separately as dopants during powder synthesis again altered the sintered density of HAp powder. It is evident from the Table V that the Hap doped with $[\text{MgCl}_2, 6\text{H}_2\text{O}]$ and TiO_2 did not show the highest sintered density. Presence of 2% ZnO in the HAp powder showed the best sintered density of 3.12 g/cc when sintered at 1250°C .

TABLE III
GREEN DENSITY AND PERCENTAGE OF THEORETICAL DENSITY

Compositions	Green density(g/cc)	% of Theoretical density
Pure HAp(calcined)	1.66	52.41%
5%Mg doped HAp (calcined)	1.69	53.36%
2%Mg doped HAp (calcined)	1.597	50.43%
5%Zn doped HAp (calcined)	1.64	51.78%
2%Zn doped HAp (calcined)	1.66	52.42%
5% Ti doped HAp (calcined)	1.65	52.01%
2%Ti doped HAp (calcined)	1.56	49.26%

TABLE IV

SINTERED DENSITY,PERCENTAGE OF DIAMETRIC,LINEAR AND VOLUMETRIC SHRINKAGE

Composition	Average Sintered Density(g/cc)	% Diame-tric Shrinkage	% Linear Shrinkage	% Volume-tric Shrinkage
Pure HAp	2.98	18.32%	18.72%	45.8%
A2.0	2.87	18.27%	18.35%	45.34%
A5.0	2.87	16.64%	16.64%	43.45%
B 2.0	3.10	18.98%	18.61%	46.54%
B5.0	2.85	17.31%	18.42%	44.15%
C2.0	2.89	18.87%	19.38%	46.94%
C5.0	2.78	16.54%	17.22%	42.31%

TABLE V

PERCENTAGE OF CONVENTIONAL HAP DENSITY OF ALL COMPOSITIONS

Composition	% of conventional HAp density
Pure HAp	93.9%
A 2.0	90.72%
A 5.0	90.52%
B 2.0	97.99%
B 5.0	90.096%
C 2.0	91.14%
C 5.0	87.82%

4.3.MECHANICAL CHARACTERIZATION

4.3.1.VICKERS HARDNESS TESTING

The average hardness of each of these composition was calculated and plotted as a function of the different indentation loads (0.3Kgf, 1 Kgf and 3 Kgf) shown in Fig 4. It is clear from the figures that the presence of Mg, Zn and Ti dopants in crystalline HAp influence its hardness.

In almost each compositions 2wt% of doping caused higher hardness values with highest load (3kgf).In other loads also hardness for 2wt% dopant concentration was

higher than that with 5%.This feature may be attributed to higher density.

Results of our hardness testing proved that the hardness of crystalline HAp ceramics in influenced by the presence of Mg, Zn and Ti as dopants during synthesis. As seen in graph of hardness in Fig 4 in 3 Kgf load we can present the Table VI.

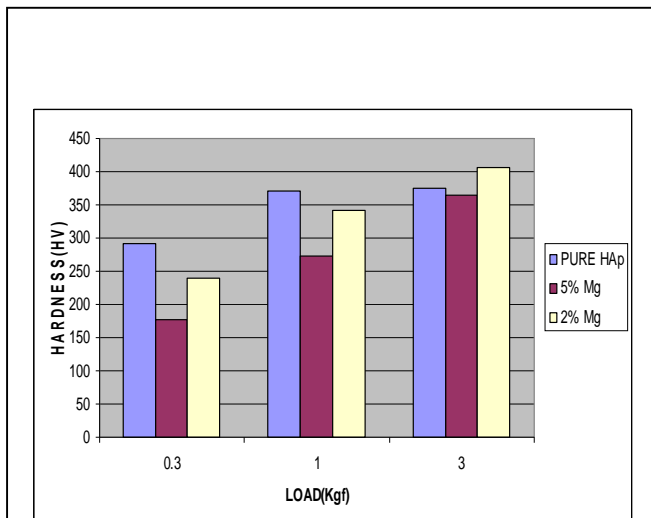


Figure 3(a).Graphical representation of Hardness of Calcined HAp and Mg doped HAp(2wt% and 5wt%)

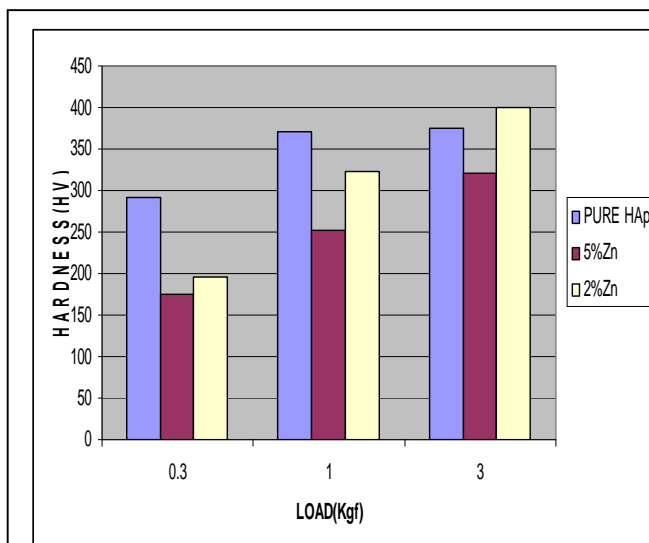


Figure 3(b).Graphical representation of Hardness of Calcined HAp and Zn doped HAp(2wt% and 5wt%)

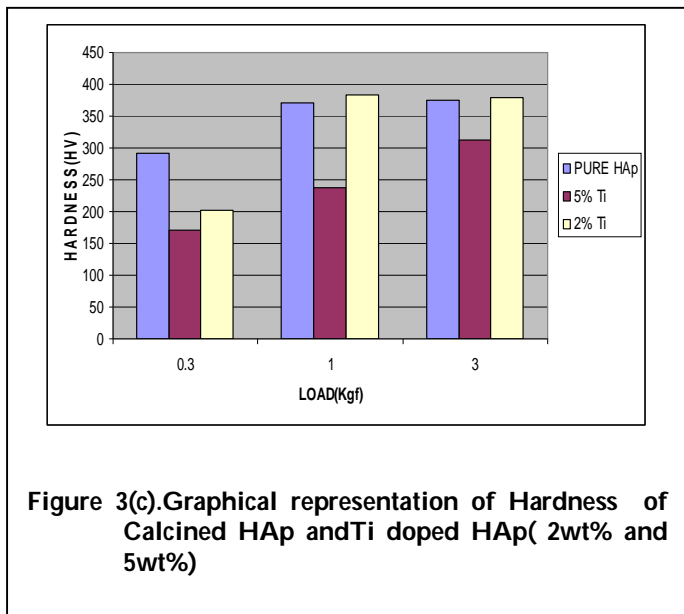


Figure 3(c). Graphical representation of Hardness of Calcined HAp and Ti doped HAp (2wt% and 5wt%)

toughness with the Inverted Optical Microscope using Image Analysis Software. The average fracture toughness and their standard deviation of each of these composition was calculated and presented in Table VII. For radial-median crack system c/a ratio (where c =average crack length, d = average diagonal) should be more than 2.5 and it was followed here.

The average crack length was $271.18\mu\text{m}$ for pure HAp, $225.05\mu\text{m}$ for 2% Mg doped HAp, $171.56\mu\text{m}$ for 5% Mg doped HAp, $210.83\mu\text{m}$ for 2% Zn doped HAp, $240.63\mu\text{m}$ for 2% Ti doped HAp and $82\mu\text{m}$ for 5% Ti doped HAp. We could not calculate the average crack length of 5% Zn doped HAp because the crack propagation was not found clearly from the site of indentation. The crack propagation path and the lateral crack growth of some specimens during the time of indentation with 3 kgf load are shown in Fig 4.

TABLE VII
AVERAGE FRACTURE TOUGHNESS AND STANDARD DEVIATION

Composition	Average Fracture Toughness (MPa $\sqrt{\text{m}}$)	Standard deviation
Pure HAp	0.4988	0.039823
A 2.0	0.63552	0.07592
A 5.0	0.9993	----
B 2.0	0.7084	0.09215
B 5.0	-----	-----
C 2.0	0.5973	0.084834
C 5.0	5.0735	0.972112

TABLE VI
AVERAGE HARDNESS AND PERCENTAGE OF HARDNESS VARIED

Composition	Average Hardness (HV)	% of Hardness increase/decrease
Pure HAp	374.31	—
A 2.0	399.42	6.71% increase
A 5.0	365.54	2.34% decrease
B 2.0	399.71	6.78% increase
B 5.0	320.593	14.35% decrease
C 2.0	378.78	1.19% increase
C 5.0	311.78	16.7% decrease

We can conclude from the above chart that percentage of hardness increased 6.71%, 6.78% and 1.19% respectively for 2 wt% of Mg, Zn and Ti dopants, but for 5 wt% of dopants of Mg, Zn and Ti % of hardness was decreased.

4.3.2.FRACTURE TOUGHNESS

Pure and doped crystalline HAp ceramics sintered at 1250°C were calculated for their fracture

We can conclude from the above chart that in most

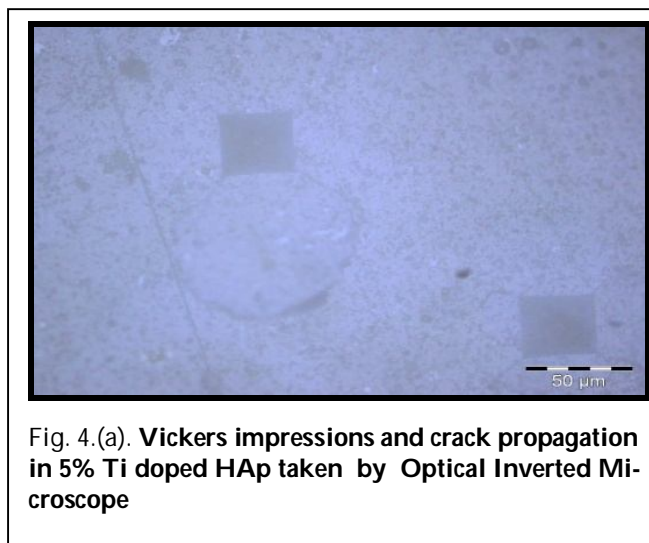


Fig. 4.(a). Vickers impressions and crack propagation in 5% Ti doped HAp taken by Optical Inverted Microscope

of the cases fracture toughness of are close to 1 or just less than 1. In the case of 5% Ti doped HAp(C 5.0) fracture toughness was highly increased. Pure HAp

increased. For 5wt% of Ti doped HAp powder fracture toughness was highly increased and has got clear impression with small cracks. In all compositions of doped HAp powder mechanical property was increased, density was high and average toughness values were increased. This is marked improvement from practical point of view compared with pure HAp powder.

6. Acknowledgment

We want to acknowledge Instrument Science and Engineering Department, Mechanical engineering Department of Jadavpur University for their technical help.

References

- [1] S.J. Kalita, S. Bose, H.L. Hosick, A. Bandyopadhyay, *Biomaterials* 25 (2004) 2331.
- [2] Yosuke Tanaka, Yoshihiro Hirata and Ryuichi Yoshinaka, *Advanced Nanostructure Material Science and Technology*, 890-0065, Japan
- [3] S.J. Kalita, D. Rokusek, S. Bose, H.L. Hosick, A. Bandyopadhyay, *J. Biomed. Mater. SRes. A* 71 (2004) 35.
- F.C. Driessens, J.W. Van Dijk, J.M. Borggreven, *Calcif. Tissue Res.* 26 (1978) 127.
- [4] R.Z. LeGeros, G. Bonel, R. Legros, *Calcif. Tissue Res.* 26 (1978) 111.
- [5] M.A. Lopes, J.D. Santos, F.J. Monteiro, J.C. Knowles, *J. Biomed. Mater. Res.* 39 (1998) 244.
- [6] R.A. Young, *J. Dent. Res.* 53 (1974) 193 (Suppl.).
- [7] S.J. Kalita, D. Rokusek, S. Bose, H.L. Hosick, A. Bandyopadhyay, *J. Biomed. Mater. SRes. A* 71 (2004) 35.
- [8] G. Georgiou, J.C. Knowles, *Biomaterials* 22 (2001) 2811.
- [9] J.D. Santos, P.L. Silva, J.C. Knowles, S. Talal, F.J. Monteiro, *J. Mater. Sci., Mater. Med.* 7 (1996) 187.
- [10] K.C.B. Yeong, J. Wang, S.C. Ng, *Biomaterials* 22 (2001) 2705.
- [11] M. Heughebaert, R.Z. LeGeros, M. Gineste, A. Guilhelm, G. Bonel, *J. Biomed. Mater. Res.* 22 (1988) 257.

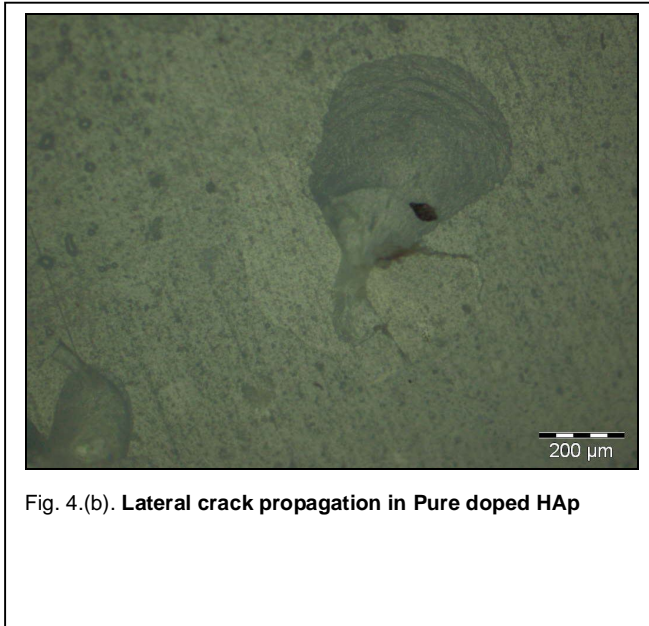


Fig. 4.(b). Lateral crack propagation in Pure doped HAp

has inherent brittleness property. We can overcome this brittleness property with these compositions. Fracture toughness is increased for 2wt% and 5 wt% of all compositions than pure HAp powder. It was also noticed 5 wt% of all compositions is better than 2 wt% of all compositions in the case of fracture toughness.

From the Fig 4(a) in case of 5% Ti doped HAp we have got very clear impression of rhombus with sharp edges. The impression resembles metal like impression with small cracks at the edges. In contrary, pure HAp impression was not so clear and at identical load (3Kgf) it suffered severe chipping due to propagation and coalescence of lateral cracks. It was also noticed that fracture toughness values of 5% Ti HAp showed large scatter. We got the lowest value of fracture toughness of 5%Ti doped HAp was $1.4359\text{MPa}\sqrt{\text{m}}$ and highest value was $9.74\text{MPa}\sqrt{\text{m}}$.

5. CONCLUSION

In our study, we prepared pure dense HAp and 2wt% and 5 wt% of Mg, Zn and Ti doped dense HAp powder using chemical route method. In XRD study, all compositions of doped and pure HAp powder we have got HAp phase. Pure HAp and all compositions of doped HAp powder we have got uniform pattern of shrinkage and almost of all cases we have got densification above 90%. Hardness was increased for 2wt% of all compositions of doped HAp powder; it may be attributed to better densification. Pure HAp is brittle material, all compositions of doped HAp powder fracture toughness was

A Chart for General and Complicated Data Visualization

M Siluvairajah

Abstract— Massive datasets arise naturally as a result of automated monitoring and transaction archival. Military intelligence data, stock trades, retail purchases, medical and scientific observations, weather monitoring, spacecraft sensor data and sensors data are all examples of data streams continuously logged and stored in extremely large volumes, which create the need for innovative data visualization solutions. Although, there are many on-going researches and developments on this field recently, there are only few solutions to visualize information for general public. In this paper, I explore different methods to use ManoStick chart to visualize information.

Index Terms—data visualization, chart, graph, manostick

1 INTRODUCTION

There is a range of visualisation tools with a range of functions, such as;

- Comparison of values with bar charts, block histograms, bubble and matrix charts
- Tag clouds to view word popularity in given text
- Data point relationships with network diagrams and scatter plots
- Parts of the whole can be visualized with pie charts and tree charts
- Track trend changes as rises and falls over time with line graphs and stack graphs
- And many more different attempts to visualize large scale information with complicated algorithms to use by researches and analysts.

However, not all are suitable to current needs and problems at hand.

Some visualization methods such as bar charts, pie charts and line graphs, are many decade old which had been created for the light-information on that time.

Some modern visualization [2] attempts such as Internet Visualizations are too complicated. According to psychologist George Miller [3], who in a seminal paper in 1956, noted the cognitive “information bottleneck” that perception and attention span impose on the amount of information people “are able to receive, process, and remember.” In this otherwise serious essay, Miller suggested that seven independent items of information, plus or minus two, might represent the limit of what most people can grasp in one moment.

According to Friedman [1], the main goal of data visuali-

zation is its ability to visualize data, communicating information clearly and effectively. It doesn't mean that data visualization needs to look boring to be functional or extremely sophisticated to look beautiful. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand, providing insights into a rather sparse and complex data set by communicating its key-aspects in a more intuitive way. Yet designers often tend to discard the balance between design and function, creating gorgeous data visualizations which fail to serve its main purpose — communicate information.

We can summarize that solutions for data visualization should be compact, based on known concept, easy to understand and able to express.

2 BACKGROUND

Introduction to ManoStick chart [4]: ManoStick chart has five microstructures as shown on figure 1 & 2, which are

1. Upper tail
2. Upper body
3. Middle body
4. Lower body
5. Lower tail

Each microstructure represents 20% of the depth or popularity, which could be quantity when working with sales, volume when working with equities etc. ManoStick chart is better suitable for variable data with depth.

To create a chart, we need to sort the variable data, find the total depth and divide the total depth with five as there are five microstructures in Mano Stick.

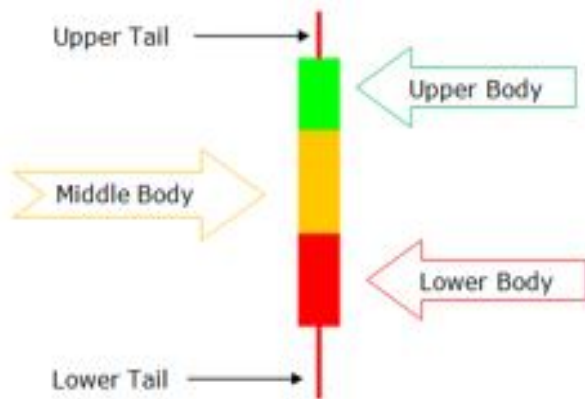


Figure 1

Now, first data set represents upper tail, the second data set represents the upper body and so on.

To differentiate microstructures, we can use different colours or patterns. Number five is convenient here as everyone used to it ever since the invention of analogue clock. When the whole divided by five, we get 20% which is each macrostructures weight.

Having four microstructures against one is rule of 80/20. Italian economist Vilfredo Pareto created a mathematical formula to describe the unequal distribution of wealth in his country, observing that 20% of the people owned 80% of the wealth. This opens up possibilities to analyse data even with single stick.

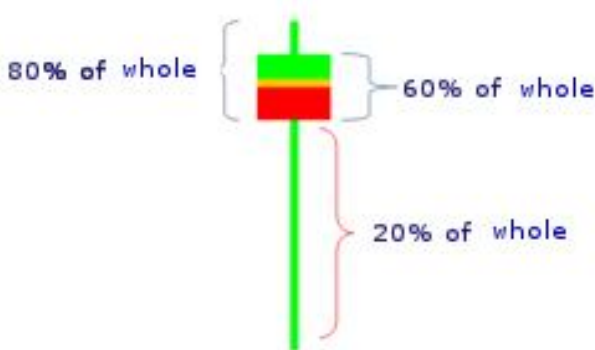


Figure 2

3. IMPLEMENTATION

What sort of data visualization might help city planners to track their population? Think of a city like Paris where thousands of people move in, move out, babies born and old people pass away. City planners are in constant challenge to offer amenities, education, health care, transport and etc. while businesses need to supply the necessary products and services.

Here, we can create a time series graph, example

monthly, using ManoStick chart. Each ManoStick is to represent the age of population for that month. By looking at the graph, we can translate lot of information easily.

We can answer representative questions such as; what is the oldest age? what is the mean average?, what age group of people are highly populated? what age group of people are least populated?

We can answer comparison questions such as; what is the difference from oldest age between two months? What age groups of people are growing from previous month? Is the average age is high or lower than the previous month?

In addition, it is possible to zoom microstructure in the form of ManoStick for deep analysis. We can continue this process as long as there are data and analysis with other soft of data such as their income in one group against the other etc.

Further, we can divide ManoStick vertically to represent to different set of data in this case, one side to represent males and the other side to represent females.

4 CONCLUSION

We have seen that ManoStick charts are able to visualize the depth, can be easily learned even by general public and take only small space to display. It can be used to represent, compare and spot trends as well as expandable with zooming. Now, it is possible to distribute more information easily to general public with a chart instead of average house price, tomorrow's temperature, cricket players batting score, children's exam result, people diagnosed with cancer or deceased people on the city last month. It opens up many more opportunities not only for statisticians but also for general public.

REFERENCES

- [1] Vitaly Friedman (2008) "Data Visualization and Info graphics" in: Graphics, Monday Inspiration, January 14th, 2008.
- [2]<http://www.webdesignerdepot.com/2009/06/50-great-examples-of-data-visualization/>
- [3] Miller, George A. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." The Psychological Review, 1956, vol. 63, pp. 81-97.
- [4] M Siluvairajah, ManoStick: An Ariel View of the Stock Market, 2009, pp.84-91 ISBN 978-0956395603
- [5] Review of Visualization in the Social Sciences: A State of the Art Survey and Report by Scott Orford, Daniel Dorling and Richard Harris, 1998.

Application of Reliability Analysis: A Technical Survey

Dr. Anju Khandelwal

Abstract— The objective of this paper is to present a survey of recent research work of high quality that deal with reliability in different fields of engineering and physical sciences. This paper covers several important areas of reliability, significant research efforts being made all over the world. The survey provides insight into past, current and future trends of reliability in different fields of Engineering, Technology and medical sciences with applications with specific problems.

Index Terms— CCN, Coherent Systems, Distributed Computing Systems, Grid Computing, Nanotechnology, Network Reliability, Reliability.

1 INTRODUCTION

THIS Traditional system-reliability measures include reliability, availability, and interval availability. Reliability is the probability that a system operates without interruption during an interval of interest under specified conditions. Reliability can be extended to include several levels of system performance. A first performance-oriented extension of reliability is to replace a single acceptable-level-of-operation by a set of performance-levels. This approach is used for evaluating network performance and reliability. The performance-level is based on metrics derived from an application-dependent performance model. For example, the performance-level might be the rate of job completion, the response time, or the number of jobs completed in a given time-interval. Availability is the probability that the system is in an operational state at the time of interest. Availability can be computed by summing the state probabilities of the operational states. Reliability is the probability that the system stays in an operational state throughout an interval. In a system without repair, reliability and availability are easily related. In a system with repair, if any repair transitions that leave failed states are deleted, making failure states absorbing states, reliability can be computed using the same methods as availability. Interval availability is the fraction of time the system spends in an operational state during an interval of interest. The mean interval availability can be computed by determining the mean time the system spends in operational states. Mean interval availability is a cumulative measure that depends on the cumulative amount of time spent in a state.

For example, how can traditional reliability assessment techniques determine the dependability of manned space vehicle designed to explore Mars, given that humanity has yet to venture that far into space? How can one determine the reliability of a nuclear weapon, given that the world has in place test-ban treaties and international

agreements? And, finally, how can one decide which artificial heart to place into a patient, given neither has ever been inside a human before? To resolve this dilemma, reliability must be: 1) reinterpreted, and then 2) quantified. Using the scientific method, researchers use evidence to determine the probability of success or failure. Therefore, reliability can be seen as an image of probability. The redefined concept of reliability incorporates auxiliary sources of data, such as expert knowledge, corporate memory, and mathematical modeling and simulation. By combining both types of data, reliability assessment is ready to enter the 21st century. Thus, reliability is a quantified measure of uncertainty about a particular type of event (or events). Reliability can also be seen as a probability.

2 APPLICATIONS OF RELIABILITY IN DIFFERENT FIELDS

Reliability is a charged word guaranteed to get attention at its mere mention. Bringing with it a host of connotations, reliability, and in particular its appraisal faces a critical dilemma at the dawn of a new century. Traditional reliability assessment consists of various real-world assessments driven by the scientific method; i.e., conducting extensive real-world tests over extensive time periods (often years) enabled scientists to determine a product's reliability under a host of specific conditions. In this 21st century, humanity's technology advances walk hand in hand with myriad testing constraints, such as political and societal principles, economic and time considerations, and lack of scientific and technology knowledge. Because of these constraints, the accuracy and efficiency of traditional methods of reliability assessment become much more questionable. Applications are the important part of research. Any theory has importance, if it is useful and applicable. Many researchers are busy these days apply-

ing concepts of Reliability in various fields of Engineering and Sciences. Some important applications are given here:

2.1 Nano-Technology

Nano-reliability measures the ability of a nano-scaled product to perform its intended functionality. At the nano scale, the physical, chemical, and biological properties of materials differ in fundamental, valuable ways from the properties of individual atoms, molecules, or bulk matter. Conventional reliability theories need to be restudied to be applied to **Nano-Engineering**. Research on Nano-Reliability is extremely important due to the fact that nano-structure components account for a high proportion of costs, and serve critical roles in newly designed products. In this paper, **Shuen-Lin Jeng et al.** [1] introduces the concepts of reliability to nano-technology; and presents the work on identifying various physical failure mechanisms of nano-structured materials and devices during fabrication process and operation. Modeling techniques of degradation, reliability functions and failure rates of nano-systems have also been discussed in this paper.

Engineer's are required to help increase reliability, while maintaining effective production schedule to produce current, and future electronics at the lowest possible cost. Without effective quality control, devices dependent on nanotechnology will experience high manufacturing costs, including transistors which could result in a disruption of the continually steady Moore's law. Nano Technology can potentially transform civilization. Realization of this potential needs a fundamental understanding of friction at the atomic scale. Furthermore, the tribological considerations of these systems are expected to be an integral aspect of the system design and will depend on the training of both existing and future scientists, and engineers in the nano scale. As nanotechnology is gradually being integrated in new product design, it is important to understand the mechanical and material properties for the sake of both scientific interest and engineering usefulness. The development of nanotechnology will lead to the introduction of new products to the public. In the modern large-scale manufacturing era, reliability issues have to be studied; and results incorporated into the design and manufacturing phases of new products. Measurement and evaluation of reliability of nano-devices is an important subject. New technology is developed to support the achievement of this task. As noted by **Keller, et al.** [2], with ongoing miniaturization from MEMS towards NEMS, there is a need for new reliability concepts making use of meso-type (micro to nano) or fully nano-mechanical approaches. Experimental verification will be the major method for validating theoretical models and simulation tools. Therefore, there is a need for developing measurement techniques which have capabilities of evaluating strain fields with very local (nano-scale) resolution.

2.2 Computer Communication Network

Network analysis is also an important approach to model real-world systems. System reliability and system unreliability are two related performance indices useful to measure the quality level of a supply-demand system. For a binary-state network without flow, the system unreliability is the probability that the system can not connect the source and the sink. Extending to a limited-flow network in the single-commodity case, the arc capacity is stochastic and the system capacity (i.e. the maximum flow) is not a fixed number. The system unreliability for $(+1)$, the probability that the upper bound of the system capacity equals can be computed in terms of upper boundary points. An upper boundary point is the maximal system state such that the system fulfills the demand. In his paper **Yi-Kuei Lin** [3] discusses about multicommodity limited-flow network (MLFN) in which multicommodity are transmitted through unreliable nodes and arcs. Nevertheless, the system capacity is not suitable to be treated as the maximal sum of the commodity because each commodity consumes the capacity differently. In this paper, Yi-Kuei Lin defines the system capacity as a demand vector if the system fulfills at most such a demand vector. The main problem of this paper is to measure the quality level of a MLFN. For this he proposes a new performance index, the probability that the upper bound of the system capacity equals the demand vector subject to the budget constraint, to evaluate the quality level of a MLFN. A branch-and-bound algorithm based on minimal cuts is also presented to generate all upper boundary points in order to compute the performance index.

In a computer network there are several reliability problems. The probabilistic events of interest are:

- * Terminal-pair connectivity
- * Tree (broadcast) connectivity
- * Multi-terminal connectivity

These reliability problems depend on the network topology, distribution of resources, operating environment, and the probability of failures of computing nodes and communication links. The computation of the reliability measures for these events requires the enumeration of all simple paths between the chosen set of nodes. The complexity of these problems, therefore, increases very rapidly with network size and topological connectivity. The reliability analysis of computer communication networks is generally based on Boolean algebra and probability theory. **Raghavendra, et al.** [4] discusses various reliability problems of computer networks including terminal-pair connectivity, tree connectivity, and multi-terminal connectivity. In his paper he also studies the dynamic computer network reliability by deriving time-dependent expressions for reliability measures assuming Markov behavior for failures and repairs. This allows computation of task and mission related measures such as mean time to first failure (MTTF) and mean time between failures (MTF).

A computer communication network (CCN) can

be represented by a set of centers and a set of communication links connecting centers which are up. The network can be represented mathematically by a graph with nodes representing centers and edges representing links. If links are simplex, then the graph is directed. Various methods have been derived to evaluate the system reliability and terminal reliability of a CCN. The methods usually involve the enumeration directly or indirectly of all the events that lead to successful communication between computer centers under consideration. However, little has been done in optimizing networks except for terminal reliability. This is because as the number of links increases, the number of possible assignments to links of the system grows faster than exponentially. **Kiu Sun-wah et al. [5]** present both mathematical and heuristic rules for optimizing the system reliability of a CCN with a fixed topology when a set of reliabilities is given. The techniques can be used to predict the system reliability for alternative topologies if the network topology is not fixed. The evaluation of the terminal reliability of a given computer communication network is a NP-hard problem. Hence, the problem of assigning reliabilities to links of a fixed computer communication network topology to optimize the system reliability is also NP-hard. Author develops a heuristic method to assign links to a given topology so that the system reliability of the network is near optimal. His method provides a way to assign reliability measures to the links of a network to increase overall reliability.

2.3 Grid Computing System

GRID computing is a newly developed technology for complex systems with large-scale resource sharing, wide-area communication, and multi-institutional collaboration. This technology attracts much attention. Many experts believe that the grid technologies will offer a second chance to fulfill the promises of the internet. The real, specific problem that underlies the Grid concept is coordinated resource sharing, and problem solving in dynamic, multi-institutional virtual organizations. Grid technology is a newly developed method for large-scale distributed systems. This technology allows effective distribution of computational tasks among different resources presented in the grid.

Yuan-Shun Dai [6], describe a grid computing systems in which the resource management systems (RMS) can divide service tasks into execution blocks (EB) and send these blocks to different resources. To provide a desired level of service reliability, the RMS can assign the same EB to several independent resources for parallel (redundant) execution. According to the optimal schedule for service task partition and distribution among resources, one can achieve the greatest possible expected service performance (i.e. least execution time), or reliability. For solving this optimization problem, the author suggests an algorithm that is based on graph theory, Bayesian approach and the evolutionary optimization approach. A virtual tree-structure model is constructed in

which failure correlation in common a communication channel is taken into account. In this reliability optimization problem, the assessment on the grid service reliability & performance is a critical component in obtaining the objective function. However, due to the size & complexity of grid service systems, the existing models for distributed systems cannot be directly applied. Thus, a virtual tree structure model was developed as the basis of the optimization model and a genetic algorithm was adapted to solve this type of optimization problem for grid task partition and distribution. In this paper author studied a case considering different numbers of resources. The Genetic Algorithm proved to be effective in accommodating various conditions including limited or insufficient resources.

Gregory Levitin et al. [7] in his paper described grid computing systems with star architectures in which the resource management system (RMS) divides service tasks into subtasks and sends the subtasks to different specialized resources for execution. To provide the desired level of service reliability, the resource management system (RMS) can assign the same subtasks to several independent resources for parallel execution. Some subtasks cannot be executed until they have received input data, which can be the result of other subtasks. This imposes precedence constraints on the order of subtask execution. Also the service reliability and performance indices are introduced and a fast numerical algorithm for their evaluation given any subtask distribution is suggested. The sharing that we are concerned with is not primarily file exchange but rather direct access to computers, software, data, and other resources. This is required by a range of collaborative problem-solving and resource-brokering strategies emerging in industry, science, and engineering. This sharing is controlled by the Resource Management System (RMS). The Open Grid Services Architecture enables the integration of services and resources across distributed, heterogeneous, dynamic virtual organizations; and also provides users a platform to easily request grid services. A grid service is desired to execute a certain task under the control of the RMS. To provide the desired level of service reliability, the RMS can assign the same subtasks to several independent resources of the same type. To evaluate the quality of service, its reliability and performance indices should be defined. Author considers the indices service reliability (probability that the service task is accomplished within a specified time), and conditional expected system time; and presents the numerical algorithm for their evaluation for arbitrary subtask distribution in a given grid with a star architecture taking into account precedence constraints on the sequence of subtask execution. Some of the very helpful practical applications are as

- Comparison of different resource management alternatives (subtask assignment to different resources),
- Making decisions aimed at service performance im-

provement based on comparison of different grid structure alternatives and

—Estimating the effect of reliability and performance variation of grid elements on service reliability and performance.

2.4 Statistical Moments / Bayes Approach

In many practical Engineering circumstances, systems reliability analysis is complicated by the fact that the failure time distributions of the constituent subsystems cannot be accurately modeled by standard distributions. **Gerard L. Reijns et al.[8]** in their paper, discuss a low-cost, compositional approach based on the use of the first four statistical moments to characterize the failure time distributions of the constituent components, subsystems and top-level system. The approach is based on the use of Pearson Distributions as an intermediate analytical vehicle, in terms of which the constituent failure time distributions are approximated. The analysis technique is presented for -out-of- systems with identical subsystems, series systems with different subsystems and systems exploiting standby redundancy. The technique consistently exhibits very good accuracy (on average, much less than 1 percent error) at very modest computing cost. In his paper, he present a low-cost, compositional approach based on the use of independent of system size. In addition, to improve their approach numeric implementation details have been outlined by him and a number of example applications from the aerospace domain have been presented in their paper.

Pandey et al. [9] provides a Bayes approach of drawing inference about the reliability of a 1-component system whose failure mechanism is simple stress-strength. The Bayes estimator of system reliability is obtained from data consisting of random samples from the stress and strength distributions, assuming each one is Weibull. The Bayes estimators of the four unknown shape and scale parameters of stress and strength distributions are also considered and these estimators are used in estimating the system reliability. The priors of the parameters of stress and strength distributions are assumed to be independent. The Bayes credibility interval of the scale and shape parameters is derived using the joint posterior of the parameters.

2.5 Genetic Algorithm

Distributed Systems (DS) have become increasingly popular in recent years. The advent of VLSI technology and low-cost microprocessors has made distributed computing economically practical. Distributed systems can provide appreciable advantages including high performance, high reliability, resource sharing and extensibility. The potential reliability improvement of a distributed system is possible because of program and data-file redundancies. To evaluate the reliability of a distributed system, including a given distribution of programs and data-files, it is important to obtain a global reliability

measure that describes the degree of system reliability. Distributed program reliability (DPR) is the probability that a given program can be run successfully and will be able to access all of the files it requires from remote sites in spite of faults occurring among the processing elements & communication links. The second measure, distributed system reliability (DSR), is defined as the probability that all of the programs in the system can be run successfully. A distributed system is a collection of processor-memory pairs connected by communication links. The reliability of a distributed system can be expressed using the distributed program reliability and distributed system reliability analysis. The computing reliability of a distributed system is an NP-hard problem. The distribution of programs and data-files can affect the system reliability. The reliability-oriented task assignment problem, which is NP-hard, is to find a task distribution such that the program reliability or system reliability is maximized. For example, efficient allocation of channels to the different cells can greatly improve the overall network throughput, in terms of the number of calls successfully supported. **Chin-Ching Chiu et al.[10]** presents a genetic algorithm-based reliability-oriented task assignment methodology (GAROTA) for computing the DTA reliability problem. The proposed algorithm uses a genetic algorithm to select a program and file assignment set that is maximal, or nearly maximal with respect to system reliability. Their numerical results show that the proposed algorithm may obtain the exact solution in most cases, and the computation time seems to be significantly shorter than that needed for the exhaustive method. The technique presented in his paper would be helpful for readers to understand the correlation between task assignment reliability and distributed system topology. **A distributed system** is defined as a system involving cooperation among several loosely coupled computers (processing elements).The system communicates (by links) over a network. **A distributed program** is defined as a program of some distributed system which requires one or more files. For a successful distributed program, the local host possesses the program, the processing elements possess the required files and the interconnecting links must be operational.

2.6 Designed Experiments

Design of experiments is a useful tool for improving the quality and reliability of products. **Designed experiments** are widely used in industries for quality improvement. A designed experiment can also be used to efficiently search over a large factor space affecting the product's performance, and identify their optimal settings in order to improve reliability. Several case studies are available in the literature.

V. Roshan Joseph et al.[11] presents the development of a integrated methodology for quality and reliability improvement when degradation data are available as the response in the experiments. The noise factors

affecting the product are classified into two groups which led to a Brownian motion model for the degradation characteristic. A simple optimization procedure for finding the best control factor setting is developed using an integrated loss function. In general, reliability improvement experiments are more difficult to conduct than the quality improvement experiments. This is mainly due to the difficulty of obtaining the data. Reliability can be defined as quality over time and therefore in reliability improvement experiments we need to study the performance of the product over time as opposed to just measuring the quality at a fixed point of time. Two types of data are usually gathered in reliability experiments: lifetime data and degradation data. **Lifetime data gives the information about the time-to-failure of the product.** In the degradation data, **a degradation characteristic is monitored throughout the life of the product.** Thus, they provide the complete history of the product's performance in contrast to a single value reported in the lifetime data. Therefore, the degradation data contain more information than the lifetime data. There are some similarities between reliability improvement, and quality improvement. Generally speaking, improving the quality will also improve the reliability. But this may not be true always. For example, suppose in a printed circuit board (PCB) manufacturing industry, tin plating is a more stable process than gold plating. Therefore in terms of improving quality the industry should prefer tin plating compared to gold plating because a better plated thickness can be achieved using tin plating. On the other hand, during customer usage, the tin will wear out faster than gold and therefore the gold-plated PCB will have higher reliability. Therefore, gold plating should be preferred for improving reliability. Thus, the choice that is good for quality need not always be good for reliability. Because of this reason, the author should find the procedure for the optimal setting of the factors considering both quality and reliability and also interaction between them.

2.7 Distributed and Coherent Systems

A Design engineer often tries to improve system reliability with a basic design to the largest extent possible subject to several constraints such as cost, weight and volume. The system reliability can be improved either by using more reliable components or by providing redundant components. If for each stage of the system, several components of different reliabilities and costs are available in the market or redundancy is allowed, then the designer faces a decision exercise which can be formulated as a nonlinear integer programming problem. **V. Rajendra Prasad et al. [12]** in their paper deals with a search method based on:

- lexicographic order and
- an upper bound on the objective function,

for solving redundancy allocation problems in coherent systems. Such problems generally belong to the class of nonlinear integer programming problems with

separable constraints and non decreasing functions. For illustration, 3 types of problems are solved using his method. A majority of problems concerning system reliability optimization are nonlinear programming problems involving integer variables. The solution methods for such problems can be categorized into:

- i) Exact methods based on dynamic programming, implicit enumeration and branch-and-bound technique
- ii) Approximate methods based on linear and nonlinear programming techniques,
- iii) Heuristic methods which yield reasonably good solutions with little computation

Each category has both advantages and disadvantages. Due to the tremendous increase in the available computing power, the exact solution deserves attention from researchers.

To derive an exact solution for a reliability optimization problem, dynamic programming can be used only for some particular structures of the objective function and constraints. It is not useful for reliability optimization of a general system, and its utility decreases with the number of constraints.

The Reliability of a Distributed Computing System is the probability that a distributed program which runs on multiple processing elements and needs to communicate with other processing elements for remote data files will be executed successfully. This reliability varies according to (1) the topology of the distributed computing system, (2) the reliability of the communication links, (3) the data files and program distribution among processing elements, and (4) the data files required to execute a program. Thus, the problem of analyzing the reliability of a distributed computing system is more complicated than the K-terminal reliability problem. In his paper, Lin et al. [13] describe several reduction methods for computing the reliability of distributed computing systems. These reduction methods can dramatically reduce the size of a distributed computing system, and therefore speed up the reliability computation.

The reliability of a distributed computing system depends on the reliability of its communication links and nodes and on the distribution of its resources, such as programs and data files. Many algorithms have been proposed for computing the reliability of distributed computing systems, but they have been applied mostly to distributed computing systems with perfect nodes. However, in real problems, nodes as well as links may fail. Min-Sheng Lin et al. [14, 15] propose in his paper, two new algorithms for computing the reliability of a distributed computing system with imperfect nodes. Algorithm I is based on a symbolic approach that includes two passes of computation. Algorithm II employs a general factoring technique on both nodes and edges. He also shows the Comparisons between both algorithms. It shows the usefulness of the proposed algorithms for computing the reliability of large distributed computing

systems.

In his paper, Raghavendra, C. S. et al. [16] present a reliability of a distributed processing system is an important design parameter that can be described in terms of the reliability of processing elements and communication links and also of the redundancy of programs and data files. The traditional terminal-pair reliability does not capture the redundancy of programs and files in a distributed system. Two reliability measures are introduced: distributed program reliability, which describes the probability of successful execution of a program requiring cooperation of several computers, and distributed system reliability, which is the probability that all the specified distributed programs for the system are operational. These two reliability measures can be extended to incorporate the effects of user sites on reliability. An efficient approach based on graph traversal is developed to evaluate the proposed reliability measures.

In his paper Deng-Jyi Chen et al. [17] Presents an algorithm for computing the reliability of distributed computing systems (DCS). The algorithm, called the Fast Reliability Evaluation Algorithm, is based on the factoring theorem employing several reliability preserving reduction techniques. The effect of file distributions, program distributions, and various topologies on reliability of the DCS is studied in brief using the Fast Reliability Evaluation algorithm. Compared with existing algorithms on various network topologies, file distributions, and program distributions, the proposed algorithm in his paper is much more economical in both time and space.

In his paper, Chiu, Chin Ching et al. [18] describe about Distributed system that provides a cost-effective means of enhancing a computer system's performance in areas such as throughput, fault-tolerance, and reliability optimization. Consequently, the reliability optimization of a distributed system has become a critical issue. A K-terminal reliability is defined as the probability that a specified set, K, of nodes is connected in a distributed system. A K-terminal reliability optimization with an order (the number of nodes in K-terminal) constraint problem is to select a K-terminal of nodes in a distributed system such that the K-terminal reliability is maximal and possesses sufficient order. It is evident that this is an NP-hard problem. This paper presents a heuristic method to reduce the computational time and the absolute error from the exact solution. The method proposed is based on not only a simple method to compute each node's weight and each link's weight, but also an effective objective function to evaluate the weight of node sets. Before appending one node to a current selected set, instead of computing the weight of all links and all nodes of each set, only the weight of a node, which is adjacent to the current selected set, and links between the node and the current selected set are accumulated. Then the proposed algorithm depends on the maximum weight to find an adequate node and assign it to the current selected set in a sequential manner until the order of K-terminal con-

straint is satisfied. Reliability computation is performed only once, thereby saving much time and the absolute error of the proposed algorithm from exact solution is very small.

2.8 Heuristic and General Reduction Methods

Distributed Computing Systems (DCS) have become a major trend in computer system design today, because of their high speed and reliable performance. Reliability is an important performance parameter in DCS design. In the reliability analysis of a DCS, the term of K-Node Reliability (KNR) is defined as the probability that all nodes in K (a subset of all processing elements) are connected. In his paper, **Ruey-Shun Chen et al. [19, 20]** proposed a simple, easily programmed heuristic method for obtaining the optimal design of a DCS in terms of maximizing reliability subject to a capacity constraint. The first half of his paper presents a heuristic algorithm which selects an optimal set of K-nodes that maximizes the KNR in a DCS subject to the capacity constraint. The second half of the paper describes a new approach that uses a K-tree disjoint reduction method to speed up the KNR evaluation. On comparing with existing algorithms on various DCS topologies, he found that the proposed algorithm is a sub-optimal design much more efficiently in terms of both execution time and space than an exact and exhaustive method for a large DCS.

In his paper, **Ruey-Shun Chen et al. [21]** present a simple, easily programmed exact method for obtaining the optimal design of a distributed computing system in terms of maximizing reliability subject to memory capacity constraints. he assume that a given amount of resources are available for linking the distributed computing system. The method is based on the partial order relation. To speed up the procedure, some rules are proposed to indicate conditions under which certain vectors in the numerical ordering that do not satisfy the capacity constraints can be skipped over. Simulation results show that the proposed algorithm requires less time and space than exhaustive method.

2.9 Tele-Communication / Neural Network

In Studies on the design of communications networks, reliability has been defined in a number of ways. **Fulya Altiparmak et al. [19]**, discusses about a probabilistic measure i.e. all-terminal reliability, is considered (this is sometimes termed overall network reliability). All-terminal reliability is the probability that a set of operational edges provides communication paths between every pair of nodes. A communications network is typically modeled as a graph with N-nodes, and L- edges; nodes represent sites (computers), and edges represent communication links. Each node, and each edge has an associated probability of failure, and the reliability of the network is the probability that the network is operational.

In his paper, **Fulya Altiparmak et al. [19]** propose a new method, based on an artificial neural network (ANN), to estimate the reliability of networks with identical link reliability. There are two significant advantages to this. The first is that a single ANN model can be used for multiple network sizes and topologies. The second advantage is that the input information to the ANN is compact, which makes the method tractable, even for large sized networks. We use the approach for networks of widely varying reliability, and then consider only highly reliable networks.

2.10 Network Reliability Optimization

Evaluation of the reliability of a network is a fundamental problem. It has application in many practical fields such as communication, digital systems, and transportation systems. The physical network is represented by a graph composed of nodes connected by directed and undirected arcs. Associated with each arc and with each node of a graph is a failure probability. **Debany, W. H. et al. [23]** provide a graph with known failure probabilities of its elements (arcs and nodes) the objective is to find the probability that at least one complete simple path (no node is visited more than once) exists between a source and a terminal node.

Developments and Improvements in information & communication technologies in recent years have resulted in increased capacities, and higher concentration of traffic in telecommunication networks. Operating failures in such high capacity networks can affect the quality of service of a large number of consumers. Consequently, the careful planning of a network's infrastructure and the detailed analysis of its reliability become increasingly important toward ensuring that consumers obtain the best service possible. One of the most basic, useful approaches to network reliability analysis is to represent the network as an undirected graph with unreliable links. The reliability of the network is usually defined as the probability that certain nodes in the graph are connected by functioning links. **Dirk et al. [24]** is discusses this Network reliability Optimization with network *planning*, where the objective is to maximize the network's reliability, subject to a fixed budget. They can develop a number of simulation techniques to address the network reliability estimation problem.

3 CONCLUSION

The behaviour of reliability is much more sensitive to change in different fields of engineering and sciences. Around twenty five papers from leading journals in different fields of reliability have been covered in this paper. This paper is reviewed various aspects of reliability research in the field of **Nano-Technology, Computer Communication Network, Grid Computing System, Statistical Moments / Bayes Approach, Genetic Algorithm** etc. We have broken down our survey into the dif-

ferent topic such as **Distributed and Coherent Systems, Network Reliability Optimization, Tele-Communication / Neural Network** etc.

4 PROPOSED FUTURE WORK

The topic reliability remains an interesting challenge for future research. The possible direction is to investigate whether reliability ranking for improving the performance of the system algorithm. Also to find the better reliability system in the sense of Distributed computing System/ Network system it is necessary to improve the time complexity for such system. In the sence of nano-technology much work is needed in particular nano-reliability field to ensure the product reliability and safety is various use condition. Some other meta-heuristic approaches besides the above may also be applicable such as Tabu-Search, Hybrid Optimization Technique and Ant Colony Optimization etc. Future aspects also exist in various communication schemes in Wireless (COBRA: Common Object Request Broker Architecture). This is also easily extensible to generic wireless network system. From the point of view of quality management, treat the system reliability as a performance index, and conduct the sensitive analysis to improve the most important component (e.g. transmission line, switch or server) will increase the system reliability most significantly. Future research can extend the problem from the single commodity case to the multicommodity case. Besides, transmission time reduction is a very important issue for an information system. Therefore researcher can extend the work to include the time attribute for each component. Reliability can also be extended for hybrid Fault-Tolerant embedded system architecture in the form of hybrid recovery Block (RB). Future research will also improved the automated controller abilities, and the human machine interface, in order to increase the efficiency of the human reasoning assistance, and to decrease the human response time.

ACKNOWLEDGMENT

The author wish to thank Prof. P. N. Tondon for the support and the encouragement which he yields to them during their research work.

REFERENCES

- [1] Shuen-Lin Jeng, Jye-Chyi Lu, and Kaibo Wang, "A Review of Reliability Research on Nanotechnology", IEEE Transactions on Reliability, Vol. 56, No. 3, Pg. 401-410, September 2007.
- [2] J. Keller, A. Gollhardt, D. Vogel, and B. Michel, "Nanoscale Deformation Measurements for Reliability Analysis of Sensors," presented at the Proceedings of the SPIE—The International Society for Optical Engineering, 2005.
- [3] Yi-Kuei Lin, "System Reliability of a Limited-Flow Network in Multicommodity Case", IEEE Transactions on Reliability, Vol. 56, No. 1, Pg. 17-25, March 2007.

- [4] Raghavendra, C. S., Kumar, V. K. P. and Hariri S., "Reliability analysis in Distributed systems", IEEE Transactions on Computers, Volume 37, Issue 3, Pg. 352 – 358, March 1988.
- [5] Kin-Sun-Wah and McAlister D.F, Reliability optimization of computer communication network, IEEE Trans. On Reliability, Vol.37, No. 2, Pp.275- 287 (1998) Dccember.
- [6] Yuan-Shun Dai, "Optimal Resource Allocation for Maximizing Performance and Reliability in Tree-Structured Grid Services", IEEE Transactions on Reliability, Vol. 56, No. 3, Pg. 444-453, September 2007.
- [7] Gregory Levitin, Yuan-Shun Dai, and Hanoch Ben-Haim, "Reliability and Performance of Star Topology Grid Service With Precedence Constraints on Subtask Execution", IEEE Transactions on Reliability, Vol. 55, No. 3, Pg. 507-515, September 2006.
- [8] Gerard L. Reijns and Arjan J. C. Van Gemund, "Reliability Analysis of Hierarchical Systems Using Statistical Moments", IEEE Transactions On Reliability, Vol. 56, No. 3, Pg 525-533, September 2007.
- [9] Pandey, M. and Upadhayay S. K., "Reliability Estimation in Stress- Strength Models: A Bayes Approach", IEEE Transactions on Reliability, December 1985.
- [10] Chin-Ching Chiu, Chung-Hsien Hsu, and Yi-Shiung Yeh, "A Genetic Algorithm for Reliability-Oriented Task Assignment With k Duplications in Distributed Systems", IEEE Transactions On Reliability, Vol. 55, No. 1, Pg. 105-117, March 2006.
- [11] V. Roshan Joseph and I-Tang Yu, "Reliability Improvement Experiments With Degradation Data", IEEE Transactions On Reliability, Vol. 55, No. 1, Pg. 149-157, March 2006.
- [12] V. Rajendra Prasad and Way Kuo, "Reliability Optimization of Coherent Systems", IEEE Transactions On Reliability, Vol. 49, No. 3, Pg. 323-330, September 2000.
- [13] M.-S. Lin and D.-J. Chen, "General Reduction Methods for the Reliability Analysis of Distributed Computing Systems", The Computer Journal, Volume 36, Issue 7, Pages 631-644, 1993.
- [14] Min-Sheng Lin, Deng-Jyi Chen and Maw-Sheng Horng, "The Reliability Analysis of Distributed Computing Systems with Imperfect Nodes", The Computer Journal, Volume 42, Issue 2, Pages 129-141, 1999.
- [15] MIN-SHENG LIN AND DENG-JYI CHEN, "GENERAL REDUCTION METHODS FOR THE RELIABILITY ANALYSIS OF DISTRIBUTED COMPUTING SYSTEMS", COMPUTER JOURNAL ISSN 00104620 CODEN CMPJA6, VOL. 36, No. 7, PP. 631-644, 1993.
- [16] Raghavendra, C. S. and Makam, S. V., "Reliability Modeling and Analysis of Computer Networks", IEEE Transactions on Reliability, Vol. R-35, No. 2, Pg. 156-160, 1986 June.
- [17] Deng-Jyi Chen and Min-Sheng Lin, "On Distributed Computing Systems Reliability Analysis Under Program Execution Constraints", IEEE Transactions on Computers, Volume 43, Issue 1, Pages 87-97, 1994.
- [18] Chiu, Chin Ching , Yeh, Yi Shiung and Chou, Jue Sam, "An Effective Algorithm for Optimal K-terminal Reliability of Distributed Systems", Malaysian Journal of Library & Information Science, 6 (2), Pg. 101-118, 2001.
- [19] Ruey-Shun Chen, Deng-Jyi Chen and Y. S. Yeh, "A New Heuristic Approach for Reliability Optimization of Distributed Computing Systems subject to Capacity Constraints", Journal of Computers & Mathematics with Applications, Volume 29, Issue 3, Pages 37-47, Feb 1995.
- [20] Ruey-Shun Chen, Deng-Jyi Chen and Y. S. Yeh, "Reliability Optimization on the Design of Distributed Computing Systems", Proceed of Int. Conf. on Computing and Information, Pages 422-437, 1994.
- [21] Ruey-Shun Chen, Deng-Jyi Chen and Y. S. Yeh, "Reliability Optimization of Distributed Computing Systems Subject to Capacity Constraints", Computers & Mathematics with Applications, Volume 29, Issue 4, Pages 93-99, February 1995.
- [22] Fulya Altıparmak, Berna Dengiz, and Alice E. Smith, "A General Neural Network Model for Estimating Telecommunications Network Reliability", IEEE Transactions on Reliability, Vol. 58, No. 1, Pg. 2-9, March 2009.
- [23] Debany, W. H. and Varshney, P. K., "Network Reliability Evaluation Using Probability Expressions", IEEE Transactions On Reliability, Vol. R-35, No. 2, Pg. 161-166, 1986 June.
- [24] Dirk P. Kroese, Kin-Ping Hui, and Sho Nariai, "Network Reliability Optimization via the Cross-Entropy Method", IEEE Transactions on Reliability, Vol. 56, No. 2, Pg. 275-287, June 2007.

About the Author: **Dr. Anju Khandelwal** is an assistant professor and HOD-Dean Academics of S.R.M.S. Women's College of Engg. & Tech. Bareilly (U.P.) Affiliated by G. B. T. U. Lucknow-India. She received her Bachelor degrees in Mathematics and Physics from Bundelkhand University Jhansi (U.P.), India in 1996, and Master degree in Mathematics with Computer Application from Bundelkhand University Jhansi (U.P.), India in 1998. She then joined the Bundelkhand University Jhansi as a lecturer under Self finance scheme. She completed his PhD degree in Operations Research from Gurukula kangri University Hardwar (Uttaranchal) in 2006. She received Master degrees in technical field that is M. Tech (Software Engineering) in 2010 from U. P. T. U. Lucknow (U.P.), Her areas of interest include parallel and distributed systems, optimization techniques, CCN and reliability Analysis. She is a life member of the IAPS.

An Innovative Quality of Service (QoS) based Service Selection for Service Orchestration in SOA

S.Neelavathi and K.Vivekanandan

Abstract— Service Oriented Architecture (SOA) has become a new software development paradigm because it provides a flexible framework that can help reduce development cost and time. SOA promises loosely coupled interoperable and composable services. Service selection in business processes is the usage of techniques in selecting and providing quality of services (QoS) to consumers in a dynamic environment. Single business process model consists of multiple service invocations forming service orchestration. It represents multiple execution paths called modeled flexibility. In certain cases, modeled flexibility can cause conflicts in service selection optimization, making it impossible to simultaneously optimize all execution paths. This paper presents an innovative approach to service selection for service orchestration that addresses this type of conflicts by encompassing status identification based availability estimation with multiple QoS constraints along with an effective quality assessment model. This model captures the expectations from the users on the multiple quality of a service and returns ratings as a feedback on the service usage. This updated rating in the service list can be used by the new user. This proposed method provides optimal services to users consistently and efficiently thereby resulting in more meaningful and reliable selection of services for service orchestration in SOA.

Index Terms— Service Oriented Architecture, Service Selection, Service orchestration, Meta-metrics, Modeled Flexibility, Rating, Multiple QoS level, local selection, global selection.

1 INTRODUCTION

Service oriented architecture (SOA) is a new paradigm for software development that promises loosely coupled, interoperable and composable components called services. Service orchestration is the execution of a single transaction that impacts one or more services in an organization. It is called as business process. Business processes are implemented by orchestrating services of different activities involved in it. Multiple QoS-based service selection results in selecting an optimal service for single activity from a set of candidate services, thereby maximizing the QoS of the entire business process. A single business process model can represent multiple execution paths known as modeled flexibility. Modelled flexibility can cause conflicts in service selection if the different optimal services are selected for the common activity in both the execution paths thereby making it impossible to optimize all execution paths. The proposed approach to service selection addresses this type of conflicts through a set of meta-metrics (probability of execution).

Status identification based availability estimation for service selection is used along with multiple QoS constraints. It is extended with an effective quality assessment model that is used to match the expectations from the user with that of rating of services held in service list. The service list is divided into four groups with all the services having the triple factor of quality rating for all the multiple QoS constraints. The feedback from the user after service usage is used to update the service list. This proposed approach of service selection with multiple QoS factors results in more meaningful and reliable selection of services used in service orchestration in service oriented architecture. Along with this, it also resolves the conflicts with modeled flexibility in business process by meta-metrics thereby ensuring selection of optimal services for all the service invocations.

2 SERVICE ORIENTED ARCHITECTURE

The service orientation and Service Oriented Architecture (SOA) are not new or revolutionary concepts. It is the next stage of evolution in the distributed computing [2]. SOA is not a Technology and is only an architectural approach. SOA is defined as "Service Oriented Architecture (SOA) is a paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains [2]." SOA includes the previously proven and successful elements from past distributed

- K.Vivekanandan is Professor in the Department of Computer Science and Engineering at Pondicherry Engineering College, Puducherry. Mobile: 9443777795 Fax: 2655101 Mail id: K.Vivekanandan@pec.edu.
- S.Neelavathi is Assistant Professor in the department of Information Technology, PKIET, Karaikal. Mobile: 9443632464. Mail Id: neelabaski@gmail.com

paradigms. These elements are combined with the design approaches to leverage recent technology in distributed computing [1].

In SOA, the loosely coupled systems do computing in terms of services. SOA separates functions into distinct units, or services, which developers make accessible over a network in order to allow users to combine and reuse them in the production of applications. These services communicate with each other by passing data from one service to another [3], or by coordinating an activity between two or more services. They use the well established standards [4].

This approach is based on the design principles of loose coupling, which is a principle by which the consumer and service are insulated from changes in underlying technology and behavior, Interoperability[2] the principle which provides the ability to support consumers and service providers that are of different programming languages on, different operating systems with different communication capabilities, Encapsulation that allows the potential consumer to be insulated from the internal technology and even the details of behavior of service, Discoverability which is used to realize the benefit of reuse [5]. Seamless integration of various systems allows data access from anywhere anytime, thereby providing services to customers and partners inside and outside the enterprise. It provides a simple scalable paradigm for organizing large networks of systems that require interoperability and develops systems that are scalable, evolvable and manageable and establishes solid foundation for business agility and adaptability.

2.1 Services In SOA

The most fundamental unit of service oriented solution logic is the service. Services in SOA comprises the below 8 distinct design principles.

■ Standardized service contract

Services express their purpose and capabilities via a service contract. It is the most fundamental part of service-orientation.

■ Service Loose Coupling

Coupling refers to the number of dependencies between modules. Loosely coupled modules have a few known dependencies whereas tightly coupled have many unknown dependencies. SOA promotes loose coupling between service consumers and providers.

■ Service Abstraction

This emphasizes the need to hide as much of the underlying details of a service as possible to preserve loosely coupled relationship.

■ Service Reusability

The same services can be reused in multiple applica-

tions. The agnostic nature of services enables them to be recombined and reused in different forms.

■ Service Autonomy

Services to carry out their functionalities consistently and reliably, needs to have a significant degree of control over its environment and resources.

■ Service Statelessness

Services are designed to remain stateful only when required. If there are stateful, then the management of excessive state information can compromise the availability of the service.

■ Service Discoverability

Service discovery is the process of discovering a service and interpretation is the process of understanding its purpose and capabilities.

■ Service Composability

It is related to its modular structure. It is composed in three ways. Application is an assembly of services, components and application logic that binds functions together. Service Federation is the collection of services managed together in a large service domain and Service orchestration is execution of single transaction that impacts one or more services in an organization.

2.2 Elements of SOA

The overall architectural model in the Fig 1 shows the elements of SOA. A service provider describes its service using Web Service Description Language (WSDL). The WSDL definition is divided into two parts: the abstract description that defines the service interface, and the concrete description that establishes the transport and location information. This WSDL definition is published to the Universal Discovery Description Interface (UDDI) service registry. SOAP (Simple Object Access Protocol) is the universally accepted standard transport protocol for messages and represents a standardized format for transporting messages. SOAP message contents are presented in message body which consists of XML formatted data. A service requestor issues one or more queries to the UDDI to locate a service and determine how to communicate with that service. Part of the WSDL provided by the service provider is passed to the service requestor. This tells the service consumer what the requests and responses are for the service provider. The service consumer uses the WSDL to send a request to the service provider. The service provider provides the expected response to the service consumer.

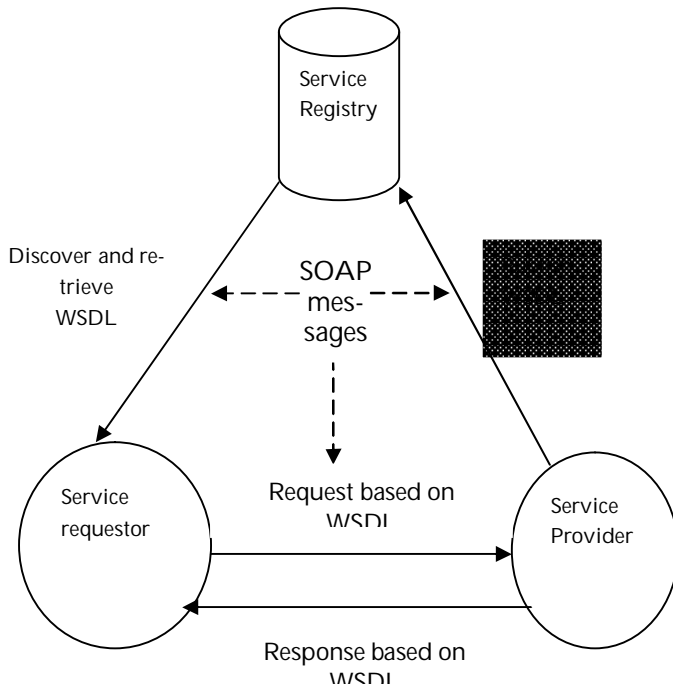


Figure 1 Elements of SOA

3 SERVICE SELECTION IN SOA

It is possible to have multiple service providers in an SOC environment who offer broadly similar services but with differing qualities. It is essential, therefore, that the selection process should select optimal service out of a group of functionality-similar services optimized for a certain property of QoS. The aim of service selection is to provide meaningful and reliable services to consumers consistently and efficiently from the set of candidate services. The figure 2 depicts a typical scenario of service selection. When the selector receives a request from a client, the selector chooses a service to invoke and returns the response message to the client. The selection process can be optimized for a certain property of QoS.

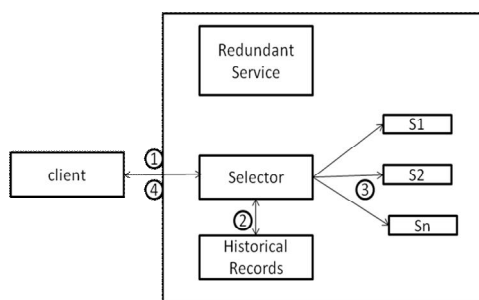


Figure 2 A Typical Scenario of Service Selection

3.1 Quality of service

Quality of service is a combination of several qualities or properties of a service such as Performance, Cost, Reliability, Availability, Reputation, and Fidelity. The performance is the time duration from a request being sent, to when the results are received. Cost refers to the amount of money that the consumer pays for using a service. The reliability is the probability that the requested service is working without a failure within a specified time frame. Availability is the quality aspect of whether the service is present or ready for immediate use. Reputation is the criterion in measuring total trustworthiness of a service and Fidelity is the average marks that are given by different consumers to the same QoS criterion.

Long term relationship between service provider and service consumer generate Service Level Agreement. The service consumer may impose several constraints to be satisfied for the services utilized thereby generating SLA. For Example the cost of service is less than 200\$, response time is less than 2ms and service should be available less than 90% of the time etc.

3.1 Classification of service selection methods

Service selection is classified as static service selection and dynamic service selection. The further is selection of optimal services from a set of services with similar functionalities during runtime while the later is selection before the runtime. Another scenario is to select services among several group of services, with each group having similar services called as Local selection and Global selection. Local Selection is to compose a system that can meet certain QoS criteria with selecting one optimal service from each group also referred as service composition whereas Global Selection does not aim at selecting an optimal service in each group, but optimizing the QoS of the composed system. Table I shows the references of static service selection and Table II shows the classification of dynamic service selection methods.

TABLE I
REFERENCES OF STATIC SERVICE SELECTION

static service selection	Pattern recognition based adaptive categorization technique and solution for services selection" IEEE2007.
	A Petri-net based specification model for Web services In Proceedings of ICWS 2004,IEEE
	Automatic Web service composition based on graph network analysis metrics Internet Systems 2005 Springer, 2005.

TABLE II
REFERENCES OF DYNAMIC SERVICE SELECTION

TABLE II
REFERENCES OF DYNAMIC SERVICE SELECTION

Based on QoS This selection process selects optimal service out of a group of functionality-similar services optimized for a certain property of QoS.	Under no global constraint	Roland Ukor, Andy Carpenter , <i>Flexible Service Selection Optimization Using Meta-metrics</i> Congress on Services-I, IEEE 2009 V.Deora,j.Shao,w.A.Gray, <i>supporting qos based selection in service oriented architecture</i> proceedings of the international conference on next generation web services practices , IEEE 2006 Y wang, J Yang, <i>Relation Based Service Networks for reliable service selection</i> proceedings of the conference on commerce and enterprise computing,IEEE 2009
	Under single global constraint	Canfora, G., Di Penta, M., Esposito, R., and Villani, M. L., " <i>An Approach for QoS-aware Service Composition based on Genetic Algorithms</i> ", Proc. of the 2005 Conf. on Genetic andevolutionary computation, ACM Press, New York, 2005. D.Liu, Z.Shao,C.Yu, <i>A heuristic Qos-aware service selection approach to web service composition</i> , International Conference on computer and Information science IEEE 2009 Lingshuang S,Lu Z, et al. <i>Dynamic Availability Estimation For Service Selection Based On Status Identification</i> , IEEE International Conference on web Services, 2008 IEEE
	Under multiple global constraint	Bang y, chi-Hung,et al. <i>Service selection model based on QoS reference vector</i> , Congress on services , IEEE 2007 <i>D.A.Menasce et al. On optimal service selection in SOA Performance Evaluation 67 (2010) 659-675</i> <i>V.Diamadopoulau et al.Techniques to support Web Service selection</i> Journal of Network and Computer Applications (2008) D.Liu, Z.Shao,C.Yu, <i>A heuristic Qos-aware service selection approach to web service composition</i> , International Conference on computer and Information science IEEE 2009
Based on Semantic web	Achieves the similarity comparison by calculating the semantic distance by QOS and context.	Z Guoping, Z Huijuan, Wang Z <i>An Approach to QoS-aware service selection in Dynamic Web service composition</i> IEEE(2007) V.X.Tran et al. <i>QoS ontology and its QoS-based ranking algorithm for Web services</i> Simulation Modelling Practice and Theory 17 (2009)
Based on improving protocol or language	Add new actions to standard UDDI to achieve dynamic UDDI process, or design a selecting language like SQL, select Web services by setting restrictive conditions.	Balke W.T, Wagner M, Kim S.M, et (2004) B.Jeong et al. <i>On the functional quality of service (FQoS) to discover and compose interoperable web services</i> Expert Systems with Applications (2009)
Based on user preference	Through users' scores on web services, achieves dynamic updating of Web services selection system, thus forming a dynamic selection process with self-evaluation function.	O.Minhyuk, B.Jongmoon,et.al, <i>An efficient approach for QoS-aware service selection based on a tree-based algorithm</i> Seventh IEEE international conference on computer and Information science 2008 <i>TQOS for automatic web service composition</i> IEEE transactions on services computing (2010)

4 PROPOSED QOS BASED SERVICE SELECTION FOR SERVICE ORCHESTRATION

Service orchestration is the execution of a single transaction that impacts one or more services in an organization, called a business process. In order to maximize the benefits of SOA, service selection is important especially in terms of providing quality of services (QoS) to consumers in a dynamic environment. The proposed work is intended to develop a more effective, meaningful and robust service selection methodology in service orchestration for conflict resolving using Meta metrics. It aims at selecting reliable and optimal services by using more than one relevant QoS category along with an effective quality assessment model.

4.1 Business Process

Business Process consists of multiple activities, each of which is a service invocation. If any of the service invocation fails, the entire transaction should be rolled back to the state that existed before execution of the transaction. Let P be a business process with a set of activities $A = \{a_1, a_2, \dots, a_n\}$. For each activity $a_i \in A$, there exists a set of candidate services $S_i = \{s_1, s_2, \dots, s_n\}$ where any of the optimal service can be used for the activity. A candidate service is selected based on multiple QoS metrics such as cost, performance, reliability, reputation and fidelity. A solution to the service selection problem is represented as a pair $\{(a_i, s_i)\}$ which assigns a service $s_i \in S_i$ for each activity a_i . The Figure 3 shows a business process with four activities. The optimal service is selected for the business dynamically on an instance-by-instance basis.

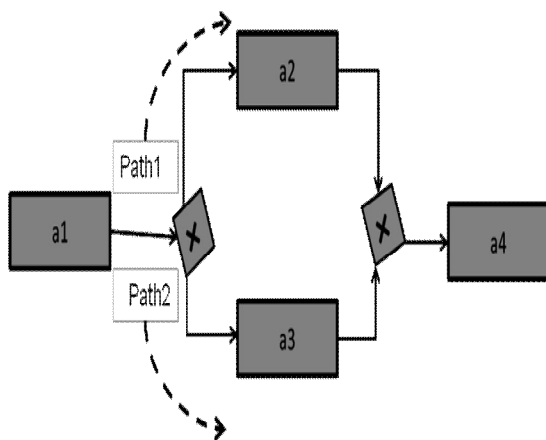


Figure 3 Example Business process

A single business process model can represent multiple execution paths, a phenomenon called modeled flexibility. An optimal solution for path1 may select candidate s_{12} for a_1 , while the optimal solution for path2 selects a different candidate s_{11} for a_1 . Therefore, it is im-

possible to obtain a solution that is simultaneously optimal for both paths, resulting in problems with service selection. This raises conflicts with modeled flexibility.

In order, to resolve these conflicts with multiple execution paths in business process, a set of process metrics called meta-metrics is used. They exist solely for the purpose of biasing the evaluation of QoS metrics for candidate services based on priorities. With probability of execution as meta-metric, activities with a higher probability of being executed are given priority to resolve the selection conflicts between the optimal solutions for two or more execution paths.

4.2 STATUS IDENTIFICATION BASED AVAILABILITY ESTIMATION FOR SERVICE SELECTION (SIBE)

The architecture of SIBE is shown in Figure 4. Quality of service assessment model is used to capture the rating from the user for the service invoked. The rating is a triple factor consisting of the expectation from the user for all the five QoS criteria's namely cost, performance, reliability, reputation and fidelity, along with the provision of actual perceived rating and quality rating offered by the user after using the service. When the user request is made for a service in the form of triple factor, the selector tries to pick up an optimal service from the service list.

The selector maintains four lists of services. Each list represents a separate status stable up, transient down, short term down and long term down. First the selector searches for the optimal service in the stable up list. If the stable up list is empty, the selector tries the transient down list. Finally the selector searches the short term down list and still further the long term list. Whenever one service status has been identified, the service is inserted into the appropriate list. The service in the service list is presented along with expectation, actual perceived value and feedback rating for all the QoS attributes from the set of users. Availability which indicates the probability of successful invocation of a service is a key property used in service selection. Multiple QoS criteria such as cost, performance, reliability, reputation and fidelity of the service are taken into consideration.

When the selection process ends, the selector puts the invocation records into historical records and reevaluates the status of the invoked services. The fragmentor algorithm takes invocation records for each service from the historical records and fragments it into equal number of fragments by randomly chosen fragment length. Each fragmented segment consists of results of service invocations in terms of either success or failure. The categorizer uses exponentially weighted average to categorize the services into appropriate list in the service list which is in turn used by the selector to select the optimal service based on user expectation.

The user provides a feedback on the service usage in terms of quality rating for all the QoS criteria's. The service list is updated with the quality rating offered by the

recent user. The new service user receives changes from the service list and uses this updated rating for reselecting their optimal services. As the inclusion of rating of many new users will subsequently incur more memory space, provision is made to remove the older entry thereby providing an efficient method of data storage.

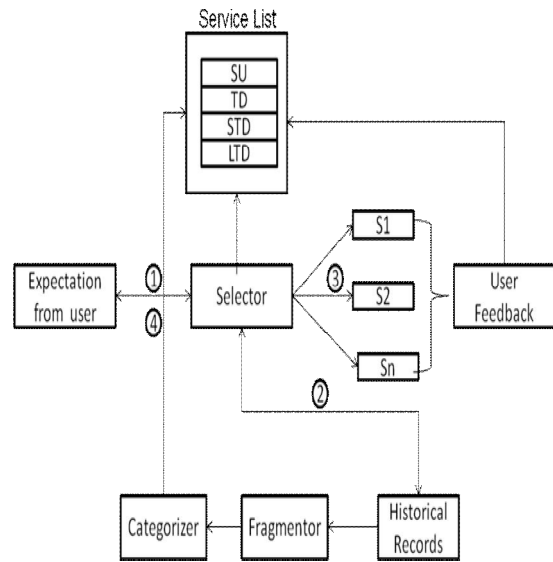


Figure 4 Architecture of SIBE

The figure 5 shows the quality of service assessment model used in SIBE. It captures the rating from the users. The user apart from mentioning the functionalities of the required service is also prompted to specify quality rating for all the QoS criteria's of the intended service. Availability is one of the key QoS attribute. Other QoS attributes considered are cost, performance, reliability, reputation and fidelity. Users specify quality rating in the form of triple factor comprising Expectation from the service, perceived value from the user and the actual rating offered by the user after using the service. The triple factor for cost is denoted as $E(c), P(c)$ and $R(c)$. Similarly it is $E(per)P(per)R(per)$ for performance, $E(r)P(r)R(r)$ for reliability, $E(rep)P(rep)R(rep)$ for reputation and $E(F)P(F)R(F)$ for fidelity.

The service held in four lists consists of triple factor for all the QoS criteria's comprising expectation, perceived value and the actual rating. Among the expectation rating provided by the user for all the QoS attributes, the highest rating of a particular attribute value is taken. Then the particular attribute with highest rating is mapped against the attribute of the service in the service list with the same expectation. If the expectation matches, then the appropriate service from the service list is selected and offered to the user. Meeting the expectation for a single attribute does not mean to satisfy the other attributes to a significant extend. So, the expectation matching can be extended to a maximum number of other

attributes, thereby providing a more meaningful and reliable service as the output. The actual rating value i.e feedback provided by the user after the service usage is added into the service list. This updated rating can be used by the user with the same intentions. Further to maintain a fixed set of entries in the service list, the older entries are always removed paving way for new entries of ratings.

Service in service list

User with Expectation	Cost $E(c) P(c) R(c)$	Performance $E(p) P(p) R(p)$	Reliability $E^R P^R R^R$	Reputation $E^R P^R R^R$	Fidelity $E(F) P(F) R(F)$
U1	★ 0.3	0.3	◆ 0.3		0.5
U2	0.8	● 0.9		0.8	
U3	★ 0.3		◆ 1.0	0.2	0.5
U4		● 0.8	◆ 0.1	0.1	0.3
Average Rating	0.46	0.67	0.47	0.36	0.43

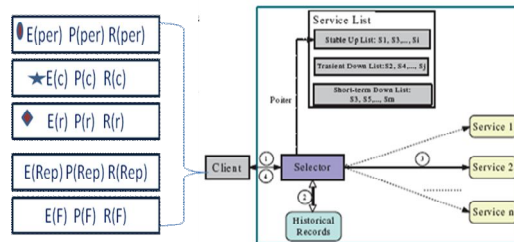


Figure 5 Quality of service assessment model used in SIBE

4.3 PROPOSED ARCHITECTURE FOR SERVICE SELECTION METHOD

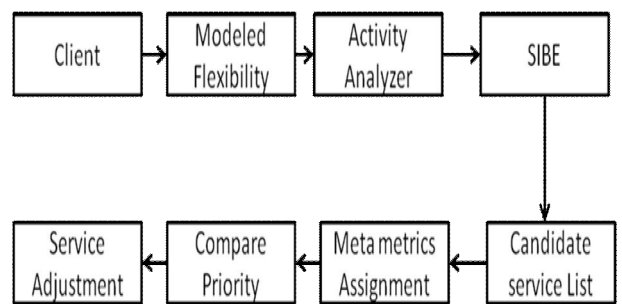


Figure 6 Proposed architecture model

The above Figure 6 shows the proposed architecture model for service selection in service orchestration of a business process model. The client in the above figure is anonymous with the business process. Every business process model represents multiple execution paths, a condition known as modeled flexibility. Activity analyzer analyzes each execution paths for the total number of ac-

tivities. For all the activities analyzed, service selection optimization is carried out dynamically on an instance-by-instance basis by SIBE (Status Identification based Availability Estimation for Service Selection) as explained above.

Each execution path represents a set of activities in a service oriented business process. Each activity in the business process involves service invocation resulting in service orchestration for the entire transaction. Service selection in a transaction can be carried out once for all activities of a process, or may be carried out on an activity-by-activity basis. Here service selection on the latter grounds is used. The optimal service selected for each activity in all execution paths is held in the candidate service list. Service orchestration is finally derived by service aggregations of services stored in candidate service list.

Meta metrics assigner check the solution set in the candidate service list comprising of activities along with their optimal services. If path1 and path 2 of business process selects same optimal service for the common activity, then there is no problem. If there is a mismatch between the services in the execution paths for the common activity, then there raises conflict of modeled flexibility. Here priority of execution of activities in execution paths is used as meta-metrics to resolve the conflicts. Comparison of priority among the activities in different execution paths is carried out by analyzing the actual working environment. Then the activity with a higher probability of execution is assigned the optimal service and the same optimal service is also allotted to other common activities in different execution paths by the service adjuster thereby overcoming the conflicts with modeled flexibility. Thus the proposed work represents an innovative approach of service selection for service orchestration by addressing the conflicts with modeled flexibility based on meta-metrics.

5.0 EVALUATION METHOD

Evaluation Environment:

1. Test Bed
2. Actual Service (Google, Yahoo etc.)

Evaluation Metrics:

1. No of services Vs the time needed to selecting optimal service.
2. No of QoS constraints Vs Score value of service.
3. Comparing the meta metrics value Vs no of conflicts arose of selecting different service for same activities in different execution paths.

4. Rating Vs three QoS levels namely low, moderate and high.
5. Selection and invocation of services Vs proxy servers stored.

6.CONCLUSION

This paper presents an innovative approach to multiple Quality of service (QoS) based service selection SIBE (STATUS IDENTIFICATION BASED AVAILABILITY ESTIMATION FOR SERVICE SELECTION) for service orchestration in service oriented architecture. SIBE results in more meaningful and reliable selection of services used in service orchestration. Execution of single transaction that impacts one or more services in an organization is called a business process. Modelled flexibility in business processes can cause conflicts between optimal service selections for activities that are common to multiple execution paths. This approach presented in this paper addresses these conflicts by using a set of meta-metrics along with meaningful service selection.

REFERENCES

- [1] Thomas Erl, (2005) *Service-Oriented Architecture: Concepts, Technology, and Design*. Prentice Hall, PTR
- [2] Judith Hurwitz, Robin Bloor, Baroudi C, Marcia K. *Service Oriented Architecture for Dummies* Wiley Publishing, Inc.
- [3] *Soa Principles Of Service Design*
- [4] Canfora, G., Di Penta, M., Esposito, R., and Villani, M L., "An Approach for QoS-aware Service Composition based on Genetic Algorithms", Proc. of the 2005 Conf. on Genetic and evolutionary computation, ACM Press, New York, 2005.
- [5] D. T. Tsesmetzis, I. G. Roussaki, I. V. Papaioannou, et al. "QoS awareness support in Web-Service semantics," in Proceedings of AICT/ICIW 2006, 2006.
- [6] Yu and K. Lin, "Service selection algorithms for Web services with end-to-end QoS constraints," in Proceedings CEC'04, pp. 129-136, 2004.
- [7] R. Ukor and A. Carpenter, "On modelled flexibility and service selection optimisation," in 9th Workshop on Business Process Modeling, Development and Support, vol. 335. CEUR-WS, 2008.
- [8] D. Ardagna and B. Pernici, "Global and local qos guarantee in web service selection," in Business Process Management Workshops, 2005, pp. 32-46.
- [9] O.Minhyuk, B.Jongmoon, et.al, "An efficient approach for QoS-aware service selection based on a Tree-based Algorithm", Seventh IEEE international conference on computer and Information science
- [10] Web Service Business Process Execution Language (WS BPEL), Version 2.0 - OASIS Committee Draft, 17th May 2006.
- [11] Lingshuang S, Lu Z, et.al, *Dynamic Availability Estimation For*

- Service Selection Based On Status Identification*, IEEE International Conference on web Services, 2008 IEEE
- [12] R. Berbner, M. Spahn, N. Repp, O. Heckmann, R. Steinmetz, *Heuristics for QoS-aware web service composition*, in: Proc. Int'l Conf. on Web Services, Sept. 2006.
- [13] G. Canfora, M. Di Penta, R. Esposito, M.L. Villani, *An approach for QoS-aware service composition based on genetic algorithms*, in: Proc. Genetic and Computation Conf., June 2005.
- [14] Roland Ukor, Andy Carpenter, *Flexible Service Selection Optimization Using Meta-metrics* Congress on Services-I, 2009 IEEE
- [15] D.A. Menascé, V. Dubey, *Utility-based QoS brokering in service oriented architectures*, in: IEEE 2007 Intl. Conf. Web Services, ICWS 2007, Application Services and Industry Track Salt Lake City, Utah, July 913, 2007, 422430.
- [16] D.A. Menascé, H. Ruan, H. Gomma, *QoS management in Service Oriented Architectures*, Performance Evaluation Journal 64 (7-8) (2007) 646663.
- [17] T. Yu, K.J. Lin, *Service selection algorithms for composing complex services with multiple QoS constraints*, in: Proc. of 3rd Int'l Conf. on Service Oriented Computing, Dec. 2005, pp. 130143.
- [18] M. Jaeger, G. Muhl, S. Golze, *Qos-aware composition of web services: A look at selection algorithm*, in: Proc. 2005 IEEE Intl. Conf. Web Services, ICWS05, 2005.
- [19] V. Deora, j. Shao, w. A. Gray, *supporting qos based selection in service oriented architecture* proceedings of the international conference on next generation web services practices, IEEE 2006
- [20] M.A. Serhani, R. Dssouli, A. Hafid, H. Sahraoui, *A QoS broker based architecture for efficient web service selection*, in: Proc. 2005 IEEE International Conference on Web Services, ICWS05, 2005.
- [21] D. Liu, Z. Shao, C. Yu, *A heuristic Qos-aware service selection approach to web service composition*, International Conference on computer and Information science IEEE 2009

Design of Boost Circuit for Wind Generator

N.Prasanna Raj, M.Mohanraj, Rani Thottungal

Abstract— This paper presents the design and implementation of a power converter for an autonomous wind induction generator (IG) feeding an isolated load through the PWM-based novel soft-switching interleaved boost converter. The output voltage and frequency of the wind IG is inherently variable due to random fluctuation of wind-speed variation. The interleaved boost converter composed of two shunted elementary boost conversion units and an auxiliary inductor. This converter is able to turn on both the active power switches at zero voltage to reduce their switching losses and evidently raise the conversion efficiency. Since the two parallel-operated elementary boost units are identical, operation analysis and design for the converter module becomes quite simple. A three-phase induction machine model and a three-phase rectifier-inverter model based on a-b-c reference frame are used to simulate the performance of the generation system. It can be concluded from the simulated results that the designed power converters with adequate control scheme can effectively improve the performance of output voltage and frequency of the IG feeding an isolated load.

Index Terms— wind power generator, rectifier-inverter circuit, pulse width modulation (PWM), interleaved boost converter, soft switching.



1 INTRODUCTION

THE characteristics of induction generator (IG) with an externally connected capacitor bank have been extensively explored for over 60 years since 1935 [1]. Due to the fast development of environmental protection concepts, the regenerative or renewable energy sources have been significantly and widely studied and evaluated in the whole world. The primary merits of IG over a conventional synchronous generator are brushless construction with squirrel-cage rotor, small size, without DC supply for excitation, less maintenance cost, and better transient performance. The IGs have been extensively utilized as suitable isolated power sources in small hydroelectric, tidal, and wind energy applications at the remote sites, rural areas, or developing countries [2 – 4]. The performance of IG supplying various static loads using different control schemes are studied and analysed in various papers. The control schemes proposed so far divides into three categories. The most economic method is by means of switching series and/or parallel capacitors connected to the IG stator winding or load side for voltage regulation. The primary disadvantage of this scheme is that the equivalent capacitance is changed in discrete form and the voltage cannot be effectively and linearly regulated. Though the voltage magnitude is controlled to a certain constant level, the output frequency of the controlled IG is significantly varied with the rotor speed. The second method is to modulate the absorbed reactive power of the IG whose stator windings is directly connected to the load and the reactive power compensator. Although the terminal voltage of the stator winding can be effectively controlled, the frequency variation problem due to random fluctuation of rotor speed is similar to the one of the first

method. The third method is the most effective control scheme, which can control both voltage and frequency of the IG within a specified level by using power electronic converters such as a rectifier-inverter module (AC-to-DC and DC-to-AC converters).

Boost converters are usually applied as preregulators or even integrated with the latter-stage circuits or rectifiers into single-stage circuits [5 – 9]. Most renewable power sources, such as photovoltaic power systems and fuel cells, have quite low-voltage output and require series connection or a voltage booster to provide enough voltage output. Several soft-switching techniques for dc/dc converters have been proposed. The main problem with these kinds of converters is that the voltage stresses on the power switches are too high in the resonant converters, especially for the high-input dc-voltage applications. Interleaved boost converters are applied as power-factor-correction front ends. An interleaved converter with a coupled winding is proposed to provide a lossless clamp. Additional active switches are also appended to provide soft-switching characteristics [10 – 13]. These converters are able to provide higher output power and lower output ripple. This paper proposes a soft-switching interleaved boost converter composed of two shunted elementary boost conversion units and an auxiliary inductor. This converter is able to turn on both the active power switches at zero voltage to reduce their switching losses and evidently raise the conversion efficiency [14], [15]. Since the two parallel-operated boost units are identical, operation analysis and design for the converter module becomes quite simple.

IG fed to an isolated load through the employment of a PWM based closed loop boost converter with controlled inverter is presented. The simulated performance of the

proposed control scheme is employed to design a PWM controller for the boost converter and inverter. The implementation and design of a power converter for an autonomous wind induction generator (IG) feeding an isolated load through the PWM-based boost-inverter circuit is simulated here.

2 METHODOLOGY

The generation system is designed with IG. The stator winding terminals of the IG are connected to the load through the rectifier, DC link, Boost circuit and inverter. The closed loop PWM signal generates proper PWM signals to switch the 2 power electronics devices of the Interleaved Boost Converter (IBC). The wind turbine rotates the IG. The IG generates power when the speed of the turbine is above the rated speed. The power generated from the IG is converted to DC with a diode bridge rectifier. The obtained DC voltage will not be in a pure DC signal. A filter circuit is used to filter out the ripple current and a pure DC voltage is obtained. This DC voltage is then boosted to the required DC level and then converted to three phase AC signal with IGBT which is driven by PWM signal. To regulate the AC output voltage the IBC is controlled by close loop PWM signals. A load is connected at the output of the inverter. The voltage-current equations of the studied IG in matrix form are listed as below.

$$v_{s(abc)} = R_{s(abc)}i_{s(abc)} + p\lambda_{s(abc)}$$

$$v'_{r(abc)} = R'_{r(abc)}i'_{r(abc)} + p\lambda'_{r(abc)}$$

Where $R_{s(abc)}$ and $R'_{r(abc)}$ are respectively the resistance matrices of the stator and rotor windings and $\lambda_{s(abc)}$ and $\lambda'_{r(abc)}$ are respectively the flux-linkage matrices of the stator and rotor windings [16].

The expression of the output power is
 $P_m = 0.5\rho AV^3 C_p$

- P_m : Mechanical output power of the turbine (W),
- C_p : Performance coefficient of the turbine,
- ρ : Air density (Kg/m³),
- A : Turbine swept area (m²),
- V : Wind speed (m/s).

Pulse-width modulation (PWM) is a very efficient way of providing intermediate amounts of electrical power between fully on and fully off. A simple power switch with a typical power source provides full power only when switched on. The term duty cycle describes the proportion of on time to the regular interval or period of time; a low duty cycle corresponds to low power, because the power is off for most of the time. Pulse-width modulation uses a rectangular pulse wave whose pulse width is modulated resulting in the variation of the average value of the waveform. If we consider a pulse waveform $f(t)$ with a low value y_{min} , a high value y_{max} and a duty cycle D , the average value of the waveform is given by

$$\bar{y} = \frac{1}{T} \int_0^T f(t) dt.$$

As $f(t)$ is a pulse wave, its value is y_{max} for $0 < t < D \cdot T$ and y_{min} for $D \cdot T < t < T$. The above expression then becomes

$$\begin{aligned} \bar{y} &= \frac{1}{T} \left(\int_0^{DT} y_{max} dt + \int_{DT}^T y_{min} dt \right) \\ &= \frac{D \cdot T \cdot y_{max} + T(1 - D) y_{min}}{T} \\ &= D \cdot y_{max} + (1 - D) y_{min} \end{aligned}$$

The simplest way to generate a PWM signal is the interceptive method, which requires only a sawtooth or a triangle waveform and a comparator. When the value of the reference signal is more than the modulation waveform, the PWM signal is in the high state, otherwise it is in the low state.

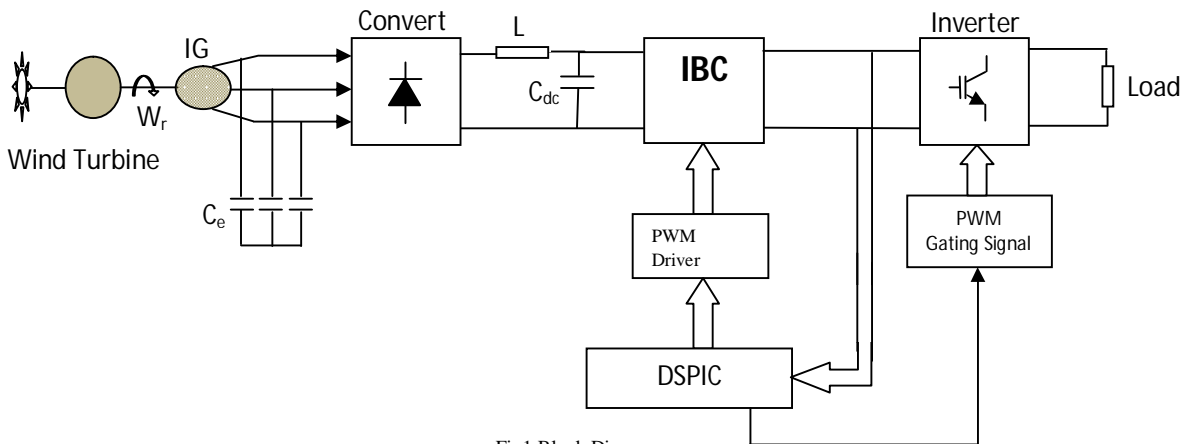


Fig1 Block Diagram

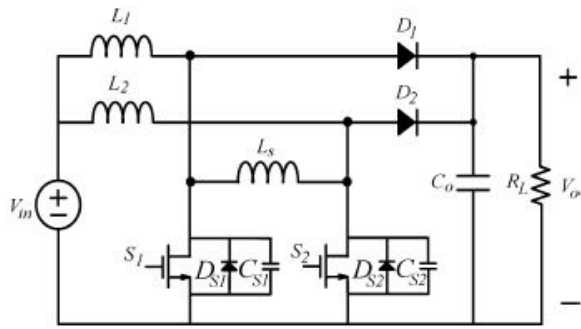


Fig 2 Interleaved Boost Converter

Fig.2 shows the proposed soft-switching converter module. Inductor L1, MOSFET active switch S1 and diode D1 comprise one step-up conversion unit, while the components with subscript "2" form the other. Dsx and Csx are the intrinsic antiparallel diode and output capacitance of MOSFET Sx, respectively. The voltage source Vin, via the two paralleled converters, replenishes output capacitor

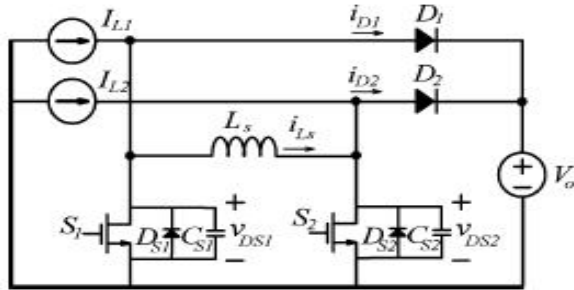


Fig 3 IBC Study

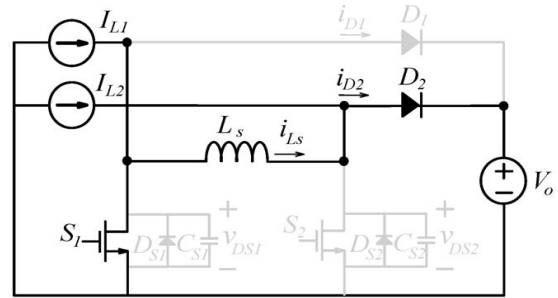
C0 and the load. Inductor Ls is shunted with the two active MOSFET switches to release the electric charge stored within the output capacitor Csx prior to the turn-ON of Sx to fulfill zero-voltage turn- ON (ZVS), and therefore, raises the converter efficiency. To simplify the analysis, L1, L2 and C0 are replaced by current and voltage sources, respectively, as shown in Fig. 3.

The proposed circuit is focused on higher power demand applications. The inductors L1 and L2 are likely to operate under continuous conduction mode (CCM); therefore, the peak inductor current can be alleviated along with less conduction losses on active switches. Under CCM operation, the inductances of L1 and L2 are related only to the current ripple specification. What dominates the output power range and ZVS operation is the inductance of Ls.

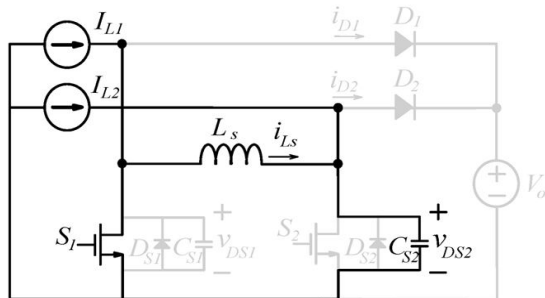
3 CIRCUIT OPERATION ANALYSIS

Before analysis on the circuit, the following assumptions are presumed.

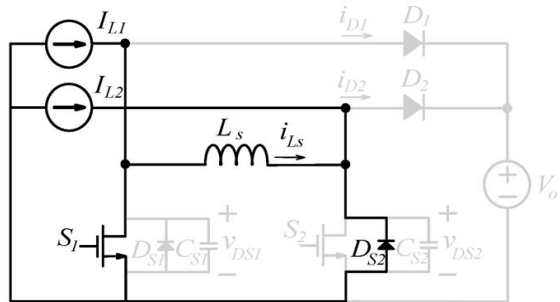
- 1) The output capacitor C0 is large enough to reasonably



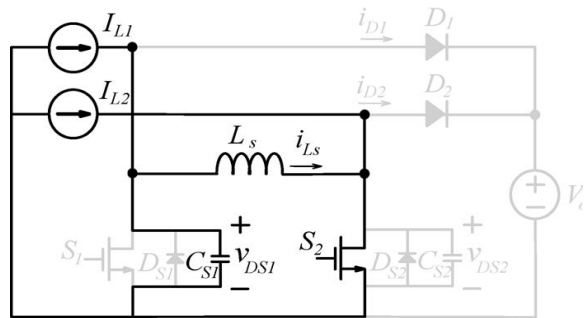
(a)



(b)



(c)



(d)

Fig 4. Modes of Operation of IBC

neglect the output voltage ripple.

2) The forward voltage drops on MOSFET S1, S2 and diodes D1 and D2 are neglected.

3) Inductors L1 and L2 have large inductance, and their currents are identical constants, i.e., $I_{L1}=I_{L2}=I_L$.

4) Output capacitances of switches Cs1 and Cs2 have the same values, i.e., $C_{s1}=C_{s2}=C_s$.

The two active switches S1 and S2 are operated with pulsewidth-modulation (PWM) control signals. They are gated with identical frequencies and duty ratios. The rising edges of the two gating signals are separated apart for half a switching cycle. The operation of the converter can be divided into eight modes, and the equivalent circuits and theoretical waveforms are illustrated in Fig. 4.

The project module is simulated using MATLAB 7.7.0(R2008b). The simulation is executed under ode23tb (stiff/TR-BDF2) state which is used to fasten the execution speed and the Zero-crossing control is disabled. The solver method is set to fast. The voltage is measured at different points in the simulation circuit. The simulated output is shown below. The system is tested with different load and wind speed. The designed system generates AC power with asynchronous generator (215HP; 400V; 50Hz). The utilizes a three-phase asynchronous machine model and a

three-phase rectifier-inverter model based on a-b-c reference frame to simulate the performance of the generation system. The IG design is made with the calculated value of resistance, flux linkage of the stator and rotor windings and with the torque equation and the number of poles. The generated power is rectified. The generated power varies with the wind speed. The power converter converts the three phase AC to DC and then the filter circuit is used for obtaining smooth DC voltage across it. This DC voltage is then boosted with the interleaved boost converter (IBC). Theoretically IBC can boost up to 200%. The DC voltage is fed to the inverter before which it is regulated to desired voltage in a closed loop with PWM technique. Thus the output at the IBC is always constant – rated voltage. This DC voltage is then regulated AC three phase with IGBT's. An IGBT inverter is designed with an open loop PWM technique and feed the AC load. PWD signal is designed in close loop and open loop system with PID and PI controller respectively. The PWM signal is generated to regulate the DC voltage that in turn gives a fixed AC supply from the inverter. Thus a regulated Three Phase AC supply is transmitted to the Load.

4 RESULTS AND DISCUSSIONS

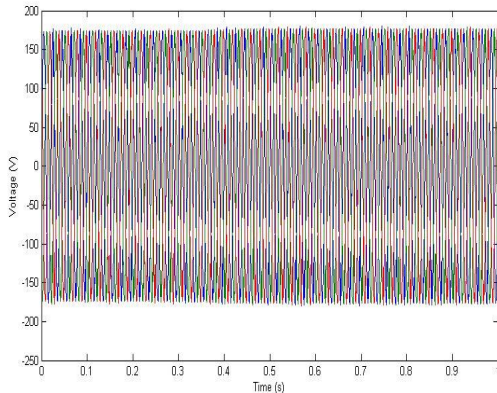


Fig 5 Input voltage 160V AC

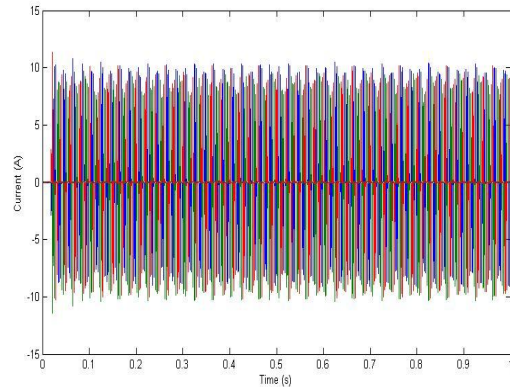


Fig 6 Input current 10A

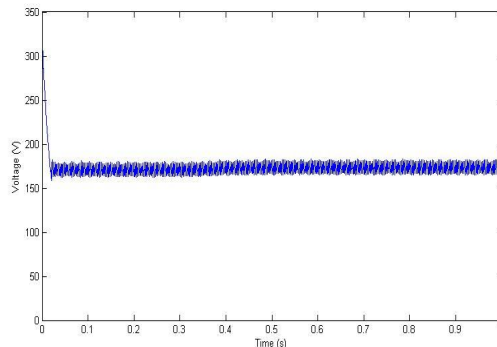


Fig 7 Rectified DC 160V

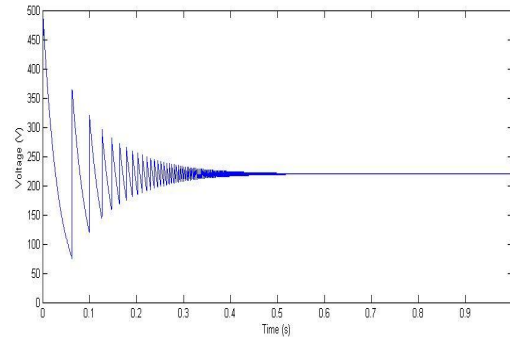


Fig 8 IBC controlled Output 220V DC

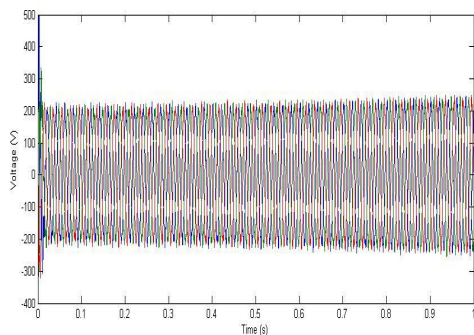


Fig 9 Output Voltage with IBC 220V AC

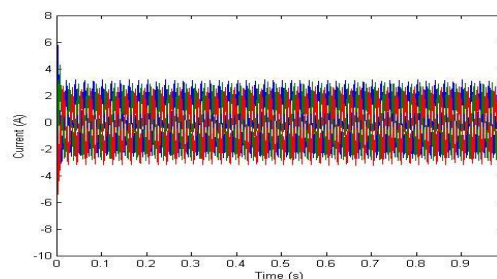


Fig 10 Output current 3A (IBC)

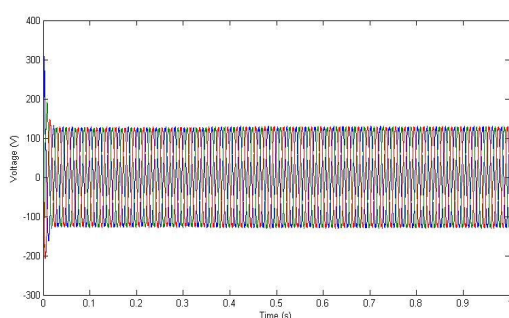


Fig 11 Output Voltage without IBC 120V

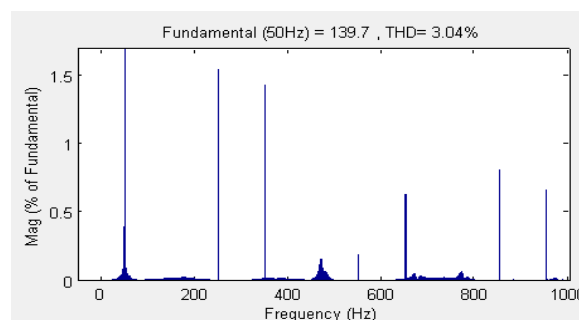


Fig 12 THD for input voltage 160V AC

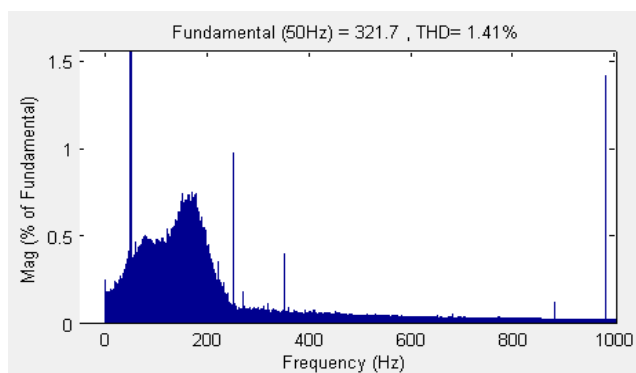


Fig 13 THD for output Voltage 220V AC

The input voltage and current generated at the wind induction generator is shown in fig 5 & 6. The generated voltage will vary for different wind speed. The generated voltage is noted for different wind velocity and at each voltage level the output voltage is noted. The examined result shows that the output voltage is maintained at 220V even when the generated voltage is below the desired voltage level. The simulated result shown here is for generated voltage of 160V and current 9A. The generated voltage is converted to DC 160V by the rectifier, shown in fig 7. This DC voltage is boosted in the closed loop circuit with the IBC (Interleaved Boost Converter) to 220V DC,

shown in fig 8. The boosted voltage is converted to Three phase AC with the inverter and the voltage obtained is the required output voltage level of 220V and the current is 3A, shown in fig 9 & 10. The output without IBC is shown in fig 11, the obtained voltage will be same the voltage generated at input. The THD value is calculated for the input voltage and the output voltage. The graph shows that the THD has decreased compared to the input voltage, shows in fig 12 & 13.

5 CONCLUSION

Thus from the above simulated results we can conclude that even at low wind speed and low power generation the required voltage can be obtained at the output. The main advantage of the system is that the minimum required wind speed for the power generation can be reduced and generator power ratings can be reduced. The gear mechanism between the turbine and the shaft can be reduced. The displayed out is taken when the wind speed is 7m/s. At this speed the generator generates 160V AC and this voltage is boosted in the IBC to 220V DC which is the required voltage level. Steady-state results under various loads show that the designed control system for IG can maintain at desired levels. The simulated results of output voltage validate the required performance of the proposed control scheme. Thus the voltage obtained at the output is regulated with the IBC and inverter block.

REFERENCES

- [1] E.D. Basset and F.M. Potter, "Capacitive excitation of induction generators," *Trans. American Institute Electrical Engineering*, vol. 54, pp. 540-545, 1935.
- [2] J. Arrilaga and D.B. Watson, "Static power conversion from self-excited induction generators," *IEE Proceedings*, vol. 125, no. 8, pp. 743-746, 1978.
- [3] R. Ramakumar, "Renewable energy sources and developing countries," *IEEE Trans. Power Apparatus and Systems*, vol. 102, no. 2, pp. 502-510, 1983
- [4] S.S. Murthy, O.P. Malik, and A.K. Tandon, "Analysis of self excited induction generators," *IEE Proceedings, Part C—Generation, Transmission, and Distribution*, vol. 129, no. 6, pp. 260-265, 1982.
- [5] B. Singh and L.B. Shilpakar, "Analysis of a novel solid state voltage regulator for a self-excited induction generators," *IEE Proceedings, Generation Transmission and Distribution*, vol. 145, no. 6, pp. 647-655, 1998.
- [6] E.G. Marra and J.A. Pomilio, "Induction-generator-based system providing regulated voltage with constant frequency," *IEEE Trans. Industrial Electronics*, vol. 47, no. 4, pp. 908-914, 2000.
- [7] E.G. Marra and J.A. Pomilio, "Self-excited induction generator controlled by a VS-PWM bidirectional converter for rural applications," *IEEE Trans. Industry Applications*, vol. 35, no. 4, pp. 877-883, 1999.
- [8] S. Wekhande and V. Agarwal, "Simple control for a wind-driven induction generator," *IEEE Industry Applications Magazine*, vol. 7, no. 2, pp. 45-53, 1999.
- [9] C. M.Wang, "A new single-phase ZCS-PWM boost rectifier with high power factor and low conduction losses," *IEEE Trans. Ind. Electron.*, vol. 53, no. 2, pp. 500-510, Apr. 2006.
- [10] Y. Jang, D. L. Dillman, and M. M. Jovanovic, "A new soft-switched PFC boost rectifier with integrated flyback converter for stand-by power," *IEEE Trans. Power Electron.*, vol. 21, no. 1, pp. 66-72, Jan. 2006.
- [11] Y. Jang, M. M. Jovanovic, K. H. Fang, and Y. M. Chang, "High-powerfactor soft-switched boost converter," *IEEE Trans. Power Electron.*, vol. 21, no. 1, pp. 98-104, Jan. 2006.
- [12] K. P. Louganski and J. S. Lai, "Current phase lead compensation in single-phase PFC boost converters with a reduced switching frequency to line frequency ratio," *IEEE Trans. Power Electron.*, vol. 22, no. 1, pp. 113-119, Jan. 2007.
- [13] K. Kobayashi, H. Matsuo, and Y. Sekine, "Novel solar-cell power supply system using a multiple-input DC-DC converter," *IEEE Trans. Ind. Electron.*, vol. 53, no. 1, pp. 281-286, Feb. 2006.
- [14] S. K. Mazumder, R. K. Burra, and K. Acharya, "A ripple-mitigating and energy-efficient fuel cell power-conditioning system," *IEEE Trans. Power Electron.*, vol. 22, no. 4, pp. 1437-1452, Jul. 2007.
- [15] G. Yao, A. Chen, and X. He, "Soft switching circuit for interleaved boost converters," *IEEE Trans. Power Electron.*, vol. 22, no. 1, pp. 80-86, Jan. 2007.
- [16] Q. Ting and B. Lehman, "Dual interleaved active-clamp forward with automatic charge balance regulation for high input voltage application," *IEEE Trans. Power Electron.*, vol. 23, no. 1, pp. 38-44, Jan. 2008.
- [17] W. Li and X. He, "ZVT interleaved boost converters for high-efficiency, high step-up DC-DC conversion," *IET Electron. Power Appl.*, vol. 1, no. 2, pp. 284-290, Mar. 2007.
- [18] W. Li and X. He, "An interleaved winding-coupled boost converter with passive lossless clamp circuits," *IEEE Trans. Power Electron.*, vol. 22, no. 4, pp. 1499-1507, Jul. 2007.

BIOGRAPHIES

N.Prasanna Raj, from Chennai, finished his UG in Coimbatore Institute of Engineering and Information Technology and pursuing his final year M.E. in Power Electronics and Drives in Kumaraguru College of Technology, Coimbatore. (praj.n@gmail.com)

M.Mohanraj, from Erode, finished his UG in Bharathiar University and he obtained PG in Power System Engineering in Annamalai University and he is currently working as Assistant Professor in EEE department, Kumaraguru College of Technology, Coimbatore and a Life member of ISTE. His research area includes Wind Energy Conversion, Solar Energy, Machines and Power Quality Issues.(mmrguru@gmail.com)

Rani Thottungal, obtained her B.E and M.E degrees from Andhra University and Doctorate from Bharathiar University. She is currently working as Professor and Head in EEE department, Kumaraguru College of Technology, Coimbatore. Her research interest includes Power System, Power Inverter and Power Quality Issues.

A Novel Method for Fingerprint Core Point Detection

Navrit Kaur Johal, Prof. Amit Kamra

Abstract- Fingerprint recognition is a method of biometric authentication that uses pattern recognition techniques based on high-resolution fingerprints images of the individual. Fingerprints have been used in forensic as well as commercial applications for identification as well as verification. Singular point detection is the most important task of fingerprint image classification operation. Two types of singular points called core and delta points are claimed to be enough to classify the fingerprints. The classification can act as an important indexing mechanism for large fingerprint databases which can reduce the query time and the computational complexity. Usually fingerprint images have noisy background and the local orientation field also changes very rapidly in the singular point area. It is difficult to locate the singular point precisely. There already exists many singular point detection algorithms, Most of them can efficiently detect the core point when the image quality is fine, but when the image quality is poor, the efficiency of the algorithm degrades rapidly. In the present work, a new method of detection and localization of core points in a fingerprint image is proposed.

Index Terms—Core Point, Delta Point, Smoothing, Orientation Field, Fingerprint Classes

1. INTRODUCTION

Fingerprints have been used as a method of identifying individuals due to the favorable characteristics such as “unchangeability” and “uniqueness” in an individual’s lifetime. In recent years, as the importance of information security is highly demanded, fingerprints are utilized for the applications related to user identification and authentication. Most Automatic Fingerprint Identification systems are based on local ridge features; ridge ending and ridge bifurcation, known as minutiae. The first scientific study of the fingerprint was made by Galton who divided fingerprint into three major classes: arches, loops, and whorls. Henry, later refined Galton’s classification by increasing the number of classification. Henry’s classification is well-known and widely accepted. Henry’s classes consist of: arch, tent arch, left loop, right loop and whorl.

At a global level the fingerprint pattern exhibits the area that ridge lines assume distinctive shapes. Such an area or region with unique pattern of curvature, bifurcation, termination is known as a singular region and is classified into core point and delta point. The singular points can be viewed as the points where the orientation field is

discontinuous.

- Navrit Kaur is pursuing M.Tech from Guru Nanak Dev Engineering College Ludhiana E-mail: navrit.johal@gmail.com
- Amit Kamra is with Guru Nanak Dev Engineering College, Ludhiana. E-mail: amit_malout@yahoo.com.

Core points are the points where the innermost ridge loops are at their steepest. Delta points are the points from which three patterns i.e. loop, delta and whorl deviate. Definitions may vary in different literatures, but this definition of singular point is the most popular one. Figure 1 below represents the core and delta points.

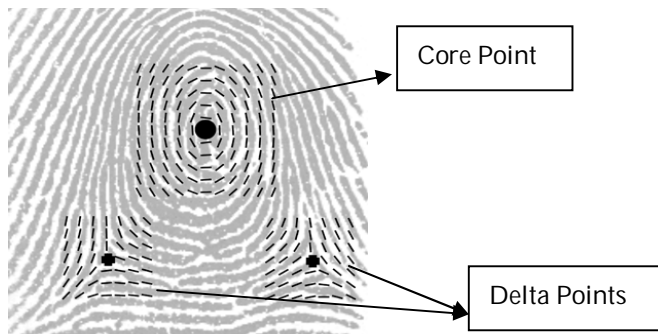


Fig 1. The Core and Delta Points on a fingerprint image

This paper is organized as follows. In section 2, are discussed the different types of fingerprints. In Section 3 is explained the drawbacks with the existing techniques of core point detection. Section 4 focuses on the problem solution. In section 5, the core point is extracted using the proposed algorithm. The experimental results performed on a variety of fingerprint images are discussed in section 6 and the conclusion and future scope is discussed in Section 7.

2 FINGERPRINT CLASSES

The positions of cores and deltas are claimed to be enough to classify the fingerprints into six categories, which include arch, tented arch, left-loop, right-loop, whorl, and twin-loop.

- Loops constitute between 60 and 70 per cent of the patterns encountered. In a loop pattern, one or more of the ridges enters on either side of the impression, recurves, touches or crosses the line of the glass running from the delta to the core, and terminates or tends to terminate on or in the direction of the side where the ridge or ridges entered. There is exactly one delta in a loop. Loops that have ridges that enter and leave from left side are called the Left Loops and loops that have ridges that enter and leave from right side are called the Right Loops. In twin loops the ridges containing the core points have their exits on different sides.
- In a whorl, some of the ridges make a turn through at least one circuit. Any fingerprint pattern which contains 2 or more deltas will be a whorl pattern
- In arch patterns, the ridges run from one side to the other of the pattern, making no backward turn. Arches come in two types, plain or tented. While the plain arch tends to flow rather easily through the pattern with no significant changes, the tented arch does make a significant change and does not have the same easy flow that the plain arch does.

In short, while classifying the fingerprints, we can make the assumption that if a pattern contains no delta then it is an arch, if it contains one (and only one) delta it will be a loop and if it contains 2 or more it will always be a whorl. If a pattern does contain more than 2 deltas it will always be an accidental whorl.

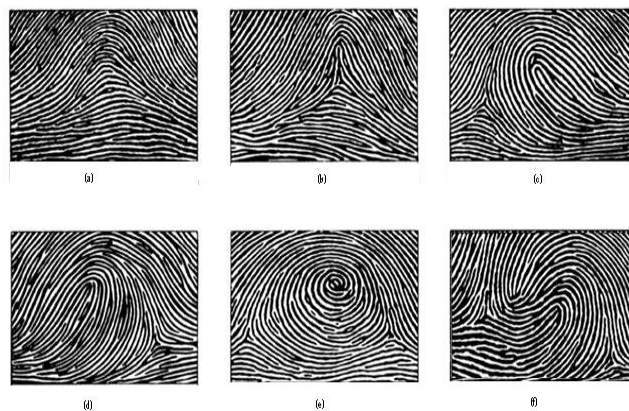


Fig 2. Classes of fingerprint (a) Arch, (b) Tented Arch, (c) Right Loop (d) Left Loop, (e) Whorl and (f) Double Loop (The double loop type is sometimes counted as whorl)

Fingerprint friction ridge details are generally described in a hierarchical order at three levels, namely, Level 1 (pattern), Level 2 (minutiae points) and Level 3 (pores and ridge shape). Automated fingerprint identification systems (AFISs) employ only Level 1 and Level 2 features. No two fingerprints are alike, but the pattern of our fingerprint is inherited from close relatives and people in our immediate family. This is considered "level 1 detail." The detail of our actual finger and palm print is not inherited. This is considered "level 2 and 3 level detail" and is used to identify fingerprints from person to person.

The following figure briefly explains the three types of levels of details in our fingerprint:

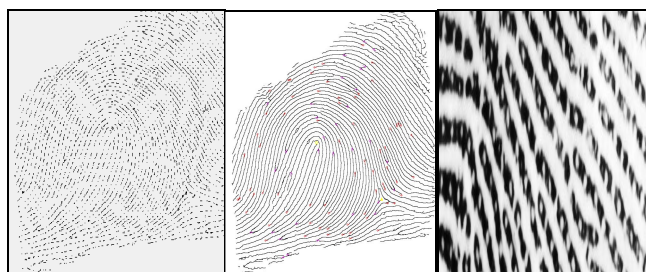


Fig 3. Level 1 , Level 2 and Level 3 Details

In this paper, we propose a new core point detection method which can precisely localize the core point and does not require further post processing as well. The proposed method only concentrates on the core point detection as most of the ridge characteristics e.g ridge endings and ridge bifurcations are present in the core block (centre).

3 PROBLEM FORMULATION

The existing techniques used for detection of core point do not produce good results for noisy images. Moreover, they

may sometimes detect spurious core point due to the inability to work efficiently for noisy images. Also techniques like Poincare Index fail for Arch type of Image. The aim of proposed algorithm is to formulate a more accurate core point determination algorithm which can produce better localization of core points avoiding any spurious points detected producing robust results for all types of fingerprints that have been discussed in this paper.

4 PROPOSED SOLUTION

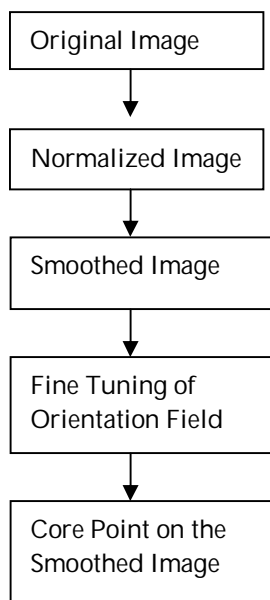


Fig 4. The proposed methodology for Core Point Detection

4.1 SEGMENTATION

The first step of the fingerprint enhancement algorithm is image segmentation. Segmentation is the process of separating the foreground regions in the image from the background regions. The foreground regions correspond to the clear fingerprint area containing the ridges and valleys, which is the area of interest. The background corresponds to the regions outside the borders of the fingerprint area, which do not contain any valid fingerprint information. Cutting or cropping out the region that does not contain valid information minimizes the number of operations on fingerprint image. The background regions of a fingerprint image generally exhibit a very low grey-scale variance value, whereas the foreground regions have a very high variance. Hence, a method based on variance threshold can be used to perform the segmentation. The steps for mean and variance based segmentation are as follows:

- a) Firstly, the image $I(i,j)$ is divided into non overlapping blocks of size $w \times w$.
- b) The mean value $M(I)$ is then calculated for each block using the following equation:

$$M(I) = \frac{1}{w^2} \sum_{i=-w/2}^{w/2} \sum_{j=-w/2}^{w/2} I(i, j) \quad (1)$$

- c) The mean value calculated above is then used to find the variance using the following equation:

$$V(I) = \frac{1}{w^2} \sum_{i=-w/2}^{w/2} \sum_{j=-w/2}^{w/2} (I(i, j) - M(I))^2 \quad (2)$$

- d) If the variance is less than the global threshold value selected empirically, then the block is assigned to be a background region; otherwise, it is assigned to be part of the foreground.

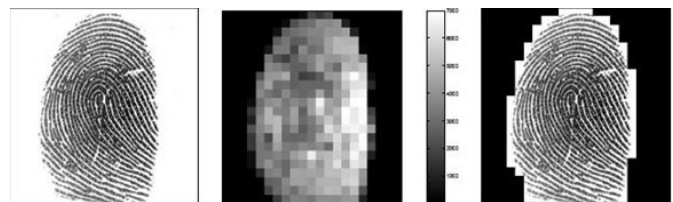


Fig 5. The result of segmentation using a variance threshold of 100 and a block size of 16×16 .

4.2 NORMALIZATION

In image processing, normalization is a process that changes the range of pixel intensity values. Normalization is sometimes called contrast stretching. In more general fields of data processing, such as digital signal processing, it is referred to as dynamic range expansion. The purpose of dynamic range expansion in the various applications is usually to bring the image, or other type of signal, into a range that is more familiar or normal to the senses, hence the term normalization. Normalization is a linear process. If the intensity range of the image is 50 to 180 and the desired range is 0 to 255 the process entails subtracting 50 from each of pixel intensity, making the range 0 to 130. Then each pixel intensity is multiplied by $255/130$, making the range 0 to 255.

Let $I(i,j)$ denote the gray-level value at pixel (i,j) , M_0 and V_0 denote the estimated mean and variance of I , respectively, and $N(i,j)$ denote the normalized gray-level value at pixel (i, j) . The normalized image is defined as follows:

$$N(i, j) = \begin{cases} M_0 + \sqrt{V_0(I(i, j) - M_i)^2} & \text{if } I(i, j) > M \\ M_0 - \sqrt{V_0(I(i, j) - M_i)^2} & \text{otherwise} \end{cases} \quad (3)$$

Normalization is a pixel-wise operation. It does not change the clarity of the ridge and valley structures. The main purpose of normalization is to reduce the variations in gray-level values along ridges and valleys, which facilitates the subsequent processing steps.



Fig 6. Result of normalization.(a) Input image(b) Normalized ($M_0 = 100$, $VAR_0 = 100$).

4.3 ORIENTATION ESTIMATION

The orientation flow is then estimated using the least square method using the following equations after dividing the input image I into non overlapping blocks of size $w \times w$ and then computing the gradients ∂_x and ∂_y at each pixel.

$$V_x(i, j) = \sum_{u=i-w/2}^{i+w/2} \sum_{v=j-w/2}^{j+w/2} 2\partial_x(u, v)\partial_y(u, v) \quad (4)$$

$$V_y(i, j) = \sum_{u=i-w/2}^{i+w/2} \sum_{v=j-w/2}^{j+w/2} \partial_x^2(u, v)\partial_y^2(u, v) \quad (5)$$

Where $\partial_x(u, v)$ and $\partial_y(u, v)$ represents gradient magnitudes at each pixel in x and y directions respectively.

The direction of block centered at pixel (i, j) is then computed using the following equation:

$$\theta(i, j) = \frac{1}{2} \tan^{-1} \left(\frac{V_y(i, j)}{V_x(i, j)} \right) \quad (6)$$

Due to the presence of noise, corrupted ridge and valley structures, minutiae etc. in the input image, the estimated local ridge orientation, $\theta(i, j)$, may not always be correct. A low-pass filter is hence used to modify the incorrect local ridge orientation.

4.4 SMOOTHING AND FINETUNING

Ridge smoothing is then performed which is a process of finding out the valid frequency of the binary image of ridges. Filters corresponding to these distinct frequencies and orientations are then generated. Fig 7 indicates the results obtained after smoothing the image.



Fig 7. Original and Smoothed Fingerprint images respectively.

The direction of gravity of progressive blocks is then determined, using the following equations ($P=3$):

$$A = \sum_{k=0}^{P-1} \sum_{l=0}^{P-1} V_x \quad \text{and} \quad B = \sum_{k=0}^{P-1} \sum_{l=0}^{P-1} V_y \quad (7)$$

As we know that singular points are the points where the orientation field is discontinuous, hence orientation plays a crucial role in estimating the core point on a fingerprint image. Hence, we need another mechanism to fine tune the orientation field so as to avoid any spurious core points and the irregularities that has occurred because of noise. The orientation field for coarse core point is then fine tuned by adjusting orientation using the following :

$$\text{if : } B_{(i,j)} \neq 0 \text{ then: } \theta = 0.5 \tan^{-1}(B/A) \quad (8)$$

else: $\theta = \pi/2$

if $\theta < 0$ then

if : $A < 0$ then: $\theta = \theta + \pi/2$

else: $\theta = \theta + \pi$

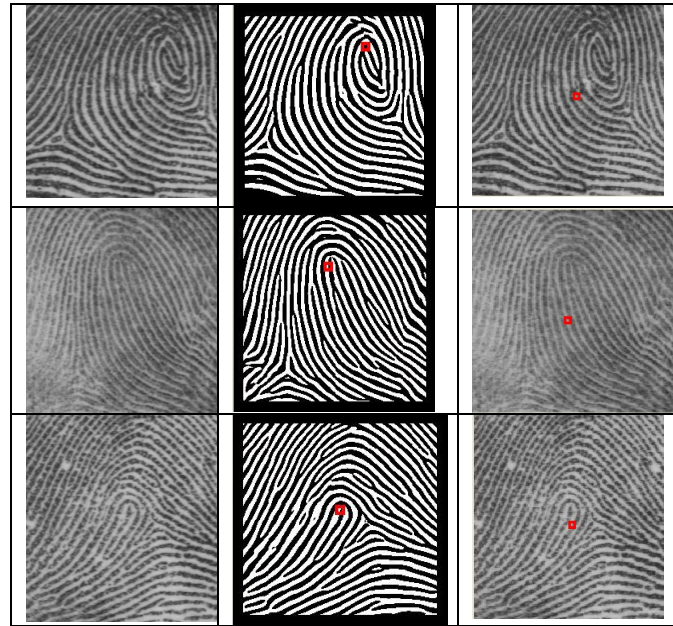
else if $A < 0$ then: $\theta = \pi/2$

Hence we calculate the value Θ , which is the orientation value of the image.

4 PROPOSED ALGORITHM

1. The original fingerprint image is first segmented and normalized using equations (1), (2) and (3).
2. Determine the x and y magnitudes of the gradients G_x and G_y at each pixel.

3. Divide the input image into blocks of size $w \times w$.
4. Estimate the local orientation using equations (4),(5),(6) and perform ridge smoothening.
5. Of the $w \times w$ image size, determine the direction of gravity of progressive blocks (non-overlapping sub block) using equation (7).
6. Fine tune the orientation field.
7. The blocks with slope values ranging to 0 to $\pi/2$ are located. Then a path is traced down until we encounter a slope that is not ranging from 0 to $\pi/2$ and that block is marked.
8. The block that has the highest number of marks will compute the slope in negative y direction and output on x and y position which will be the core point.



The proposed algorithm gives us near accurate results for noisy images as shown below:

5 RESULTS AND DISCUSSION

The proposed technique can localize the fingerprints at a good success. Experiments have been performed on nearly 180 fingerprints from different fingers and a variation in noise has been taken into consideration. The results have been obtained in MATLAB, the core point positions hence obtained are used to narrow down the search when using a huge database of fingerprint images in applications like biometrics security, forensics etc. The following table clearly compares the accuracy of results of the proposed technique with the existing state of the art technique.

TABLE 1: Comparison of results of the proposed technique with the existing technique.

Original Image	Proposed Technique	Existing Technique



Fig 10. Accuracy of the proposed technique for a poor input image.

But the algorithm failed its core point detection for images having a very poor quality as shown in figure 11. Those fingerprints are too oily or wrinkled.

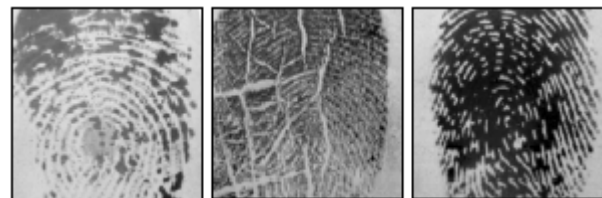


Fig 11. Poor quality fingerprint images

6 CONCLUSION AND FUTURE SCOPE

This paper proposes a novel method to consistently and precisely locate the singular points in fingerprint images. The method applied is based on the enhanced fingerprint

image orientation reliability. Our future work will focus on improvements in locating the secondary singular points of fingerprint images and classifying the fingerprints based upon the location of the singularities so that the computational time for matching fingerprints in a huge database may reduce.

7 REFERENCES

- [1] Keokanlaya Sihalath, Somsak Choomchuay, and Kazuhiko Hamamoto, "Core Point Identification with Local Enhancement", JCSSE, May 13-15, 2009 (Vol. 1)
- [2] Mohammed S. Khalil, Dzulkifli Muhammad, Muhammad K. Khan and Khaled Alghathbar, "Singular Point Detection using Fingerprint Orientation Field Reliability" , International Journal of Physical Sciences Vol. 5(4), pp. 352-357, April 2010
- [3] Wang Feng, Chen Yun , Wang Hao, Wang Xiu-you "Fingerprint Classification Based on Improved Singular Points Detection and Central Symmetrical Axis" , 2009 International Conference on Artificial Intelligence and Computational Intelligence
- [4] Filipe Magalhães, Hélder P. Oliveira, Aurélio C. Campilho, "A New Method for the Detection of Singular Points in Fingerprint Images" , 978-1-4244-5498-3/09©2009 IEEE
- [5] Galton F., Fingerprint, McMillan, London, 1892.
- [6] Henry E., "Classification and uses of fingerprints" , Rout ledge, London, 1900.
- [7] Weiwei Zhang, Yangsheng Wang , "Core-Based Structure Matching Algorithm of Fingerprint Verification" , National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
- [8] Tomohiko Ohtsuka, Daisuke Watanabe, Hiroyuki Aoki1, "Fingerprint Core And Delta Detection By Candidate Analysis" , MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, Japan
- [9] S.Chikkerur, C. Wu and Govindaraju, "A Systematic Approach for Feature Extraction in Fingerprint Images", ICBA 2004.
- [10] L. Hong, Y. Wan, and A. Jain, "Fingerprint Enhancements: Algorithm and Evaluation" , Proc. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.20, no. 8, pp. 777-789, 1998.
- [11] M. Liu, X. Jiang, and A. C. Kot, "Fingerprint Reference Point Detection" , Journal of Applied Signal Processing, vol. 4, pp. 498-509, 2005.
- [12] C.-Y. Huang, L. Liu, and D. D. Hung, "Fingerprint Analysis and Singular Point Detection".
- [13] Ramandeep Kaur, Parvinder S. Sandhu, Amit Kamra, "A Novel Method for Fingerprint Feature Extraction" 2010 International Conference on Networking and Information Technology 978-1-4244-7578-0 © 2010 IEEE
- [14] Bob Hastings, "A Statistical Investigation of Fingerprint Patterns Department of Computer Science & Software Engineering The University of Western Australia, Australia.

Navrit Kaur Johal received her B.Tech degree in Computer Science and Engineering from AIET, Faridkot and is pursuing her M.Tech in Computer Sc. and Engg from Guru Nanak Dev Engg. College, Ludhiana.

Amit Kamra received his B.Tech degree in Electronics Engineering in 2001. and completed his M.Tech in Information Technology in 2004. He is working as Associate Professor in Guru Nanak Dev Engg. College, Ludhiana and is currently pursuing his Ph.D in image processing.

Robust Impulse Eliminating Internal Model Control of Singular Systems: A Robust Control Approach

M.M. Share Pasand, H.D. Taghirad

Abstract— The problem of model based internal control of singular systems is investigated and the limitations of directly extending the control schemes for normal systems to singular ones were analyzed in this paper. A robust approach is proposed in order to establish the control scheme for singular systems, and moreover, to present a framework for robust control of singular systems in presence of modeling uncertainties. The theory is developed through a number of theorems, and several simulation examples are included and their physical inter-pretations are given to verify the proof of concept.

Index Terms— singular systems, Impulsive behavior, internal model control, Model based control, robust control, tracking problem, impulse elimination.

1. INTRODUCTION

Singular systems represent a general framework for linear systems [1]. A singular model is an appropriate model for describing large scale interconnected systems, constrained robots and other differential algebraic systems with linear algebraic constraints [2]. Also singular models can be utilized to model a system when the dependent variable is displacement rather than time [3]. Since the first time they were introduced [4], several efforts have been made to control singular systems [5-9]. As the singular systems were firstly introduced in the state space form representation [4], they were usually studied in time domain. In [5] the problem of finite mode pole placement is studied, while simultaneous impulse elimination and robust stabilization problem is considered in [6], robust Eigen-structure assignment of finite modes is studied in [7]. In [8] strict impulse elimination is studied using time derivative feedback of the states and [9] investigated the output feedback control using a compensator. In fact most of the existing methods are extensions of the control schemes for standard systems [5],[6],[10]. In the singular system control context the control objectives are more complicated due to the existing obstacles such as algebraic loop phenomenon, impulsive behavior [11], and regularity of the closed loop [8,9]. Unlike the time domain methods, there are very few works on the frequency domain control of singular systems. In the frequency domain, the tracking problem, robust control problem and impulse elimination can be treated more conveniently. Specifically the so called Internal Model Control (IMC) method provides a very interest-

ing framework for analyzing the algebraic loop, regularity of the closed loop and impulse elimination problems of singular systems. Furthermore, most of existing methods in robust control of singular systems are limited to study a special form of uncertainty. They assumed matrix E to be exactly known [6, 7 and 10]. This assumption is more restrictive than it appears, because it limits the system to be impulsive while some uncertainties may exist which lead to a strictly proper system for a singular model. Therefore this paper suggests a new concept for robust control. While previous works on robust control focuses on robust stability and robust performance, as it comes to descriptor systems, robust properness of closed loop should be studied. The internal model framework for controlling singular systems provides a more logical uncertainty model and release the restrictive assumptions made in the existing state space methods for robust control of singular systems. Also it provides offset free tracking capability of the closed loop as well as being able to well treat delayed systems. The main obstacle which arises in the internal model control of singular systems is that the internal model cannot be implemented easily, because it is generally improper. Even in computer aided control systems it is not easy to simulate a singular system, since the discrete model needs future input data to determine the system state vector at the present time [1]. This problem results in an inevitable mismatch between the plant and the parallel model used in IMC.

Notice that, defining the disk shaped multiplicative uncertainty leads to an unbounded uncertainty profile which is not suitable in robust design of singular systems. This paper provides solution to the latter problem by introducing the singular internal model filter in series with the conventional internal model filter. The aforementioned filter eases the design procedure, bounds the uncertainty profile. Also it makes the closed loop strictly proper and eliminates impulsive modes by smoothing the control action as much as needed. Another role of the introduced filter is to make it possible to design robust control-

- M.M. Share Pasand was graduated from KNT University of technology. He is currently within IAU Malard branch.
E-mail: momeshpa@gmail.com.
- Dr. H.D. Taghirad is a professor within the control systems group, KNT University of technology.

ler in the conventional context. The paper is organized as follows. In the next section backgrounds are discussed and the obstacles in control of singular systems are presented, and some major limitations of the direct extension of IMC are explained. In the third section the proposed method is studied and the filter design procedure is illustrated. In the fourth section several examples and simulations are given to examine the algorithm both in terms of robustness properties and closed loop performance. Finally, the concluding remarks are given in last section.

CONTROL OBJECTIVES IN SINGULAR CONTROL SYSTEMS

Definitions and Singular Systems Characteristics

As Descriptor models are a straight extension of standard state space models [1], control problem for these systems has a wider range of objectives. A control system for a standard plant is designed such that the closed loop is stable and has a predefined performance and acceptable robustness properties. A singular control system, on the other hand, should be designed such that it is impulse free, regular and doesn't include any algebraic loops in addition to the aforementioned properties. These control objectives combined with the standard objectives make the control of singular systems more challenging. Robust control of singular systems is the most challenging issue because it requires robustness not only in the stability and performance but also in regularity and properness. State space robust control schemes require robust observers in order to work properly and do not guarantee strict properness of the closed loop also they usually result in more complicated derivation algorithms. The main advantage to use internal model control scheme in here, is that the IMC provides an effective tool in frequency domain without introducing complicated methods in evaluation of closed loop performance and stability. Therefore, IMC can be regarded as a proper alternative for existing state space methods. Also IMC provides a simple framework for algebraic loop and properness analysis of singular control systems which is much simpler than that in state space methods or other frequency domain schemes. Consider the following state space description:

$$\begin{aligned} E\dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \tag{1}$$

Definition1: System (1) is impulse free if and only if:

$$\deg|sE - A| = \text{rank } E \tag{2}$$

The nullity index of E is called singularity index of a singular system (1) in this paper.

Remark1: Note that the following general inequality always holds:

$$\deg|sE - A| \leq \text{rank}(E) \tag{3}$$

Corollary1: A singular system is called impulse free, if and only if, it doesn't exhibit impulses in its impulse response.

Definition2: A singular system is called minimal if it is observable and controllable. The minimality of the plant is presumed throughout this paper.

Definition3: A transfer function is strictly proper, bi-proper and improper if the following limit is zero, a finite nonzero value and infinite respectively. Strictly proper and bi-proper systems may be generally named as proper.

$$\alpha = \lim_{s \rightarrow \infty} \eta(s)$$

Lemma1: A state space realization of a singular system is impulse free if and only if its transfer function has a nonnegative relative degree i.e. it is proper.

Proof: The transfer function matrix from input to state for system (1) can be computed as:

$$\eta(s) = C(sE - A)^{-1}B = \frac{C \cdot \text{adj}(sE - A)B}{|sE - A|}$$

It is known that degree of the nominator is equal to rank of E at most, therefore if condition (2) is satisfied then the system transfer matrix will be proper and if not, the transfer matrix may be improper. On the other hand if transfer matrix is not proper condition (2) is not satisfied. ■

Remark2: The observability assumption is essential for the above lemma. It can be shown that it may be a number of unobservable impulsive modes which do not appear in the output. Also note that condition (2) is a general condition for impulse free systems but in order to equate it with corollary 1, the observability assumption is needed.

Lemma2: In the unity output feedback structure the closed loop system is strictly proper if the compensator/plant combination is strictly proper.

Proof: Expand the nominator and denominator by their respective Taylor series.

$$\eta(s) = \frac{\sum_{k=1}^{\infty} a_k s^{-k}}{1 + \sum_{j=1}^{\infty} b_j s^{-j}}$$

Because CP is supposed to be strictly proper, the largest term in its expansion has a negative power therefore the denominator has a greater degree than the nominator and thus the closed loop system is strictly proper. ■

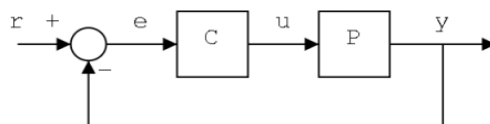


Figure1: Feedback structure

Remark3: Note that Lemma2 provides a sufficient condition. The necessary and sufficient condition is derived later. Lemma2 shows that why the objective of properness has not been considered before the introduction of descriptor systems. Assuming strictly proper functions for plant and compensator, it is trivial that the closed loop system is strictly proper. Also for a strictly proper plant and a bi-proper compensator the closed loop will be bi-proper.

Lemma3: For a bi-proper plant and a bi-proper compensator, the closed loop will be improper if and only if:

$$\lim_{s \rightarrow \infty} CP = -1$$

Lemma4: In order to have a strictly proper closed loop system with a unit feedback, if the plant is improper the compensator should be strictly proper with a sufficiently large relative degree.

Proof: For the closed loop to be strictly proper, the compensator/plant should be strictly proper according to lemma2, therefore the compensator should be strictly proper. ■

Robust Internal Model Control of Singular Systems

In IMC structure the parallel model or internal model is inevitably proper or strictly proper. Therefore there always exists a mismatch between the plant and model. For a continuous output especially in case of initial jumps of the input, it is required that the plant and the internal model have the same infinite gain and the compensator is strictly proper. This issue can be treated by a smoothing pre-filter for reference signal but the method is not robust against model uncertainties. The IMC filter is conventionally used to enhance robustness properties by victimizing closed loop response and making the compensator implementable (proper). Moreover, it accounts for online adaptation of the control system by adjusting the filter time constant. In this paper we extend this approach by using a second IMC filter which assures the closed loop to be strictly proper and has a smooth response by compensating the singular plant impulsive behavior. The singular internal model control filter or SIMC filter is designed to yield a continuous smooth response and a robust IMC design for singular systems. In fact by using a parallel strictly proper model in IMC, the uncertainty will become unbounded and the robust control will not be feasible any more. Therefore the SIMC filter has another role of bounding the uncertainty profile and making the robust control problem feasible. The disk-type uncertainty profile is usually assumed in robust control schemes, is described by the following relation.

$$\frac{|p(j\omega) - \tilde{p}(j\omega)|}{|\tilde{p}(j\omega)|} \equiv l_m(\omega) \leq \bar{l}_m(\omega) \quad (4)$$

This uncertainty description allows us to incorporate several singular systems in the design while the state space un-

certainty descriptions are limited to represent only singular systems with a pre-specified singularity index. If one augments the improper plant by high frequency stable poles a strictly proper model can be obtained, which has a very close behavior to plant at least at low enough frequency range. Larger poles results in closer response to that of the plant in wider bandwidths. However, in this way the uncertainty becomes unbounded. In particular assume a polynomial of stable real poles with a unit steady state gain namely D , then one can write:

$$\tilde{p}(s) = \frac{p(s)}{D(s)} \quad (5)$$

The above description for model is the most natural selection for a strictly proper model, whose behavior is as close as possible to that of the plant. However, in this situation the mismatch between plant and model is not included in a disk shaped region. In other words the uncertainty bound will be infinity. Now we can take different approaches: Choose another internal model which yields bounded uncertainty; developing new theory for this kind of uncertainty; or modify the plant input in order to bind the uncertainty as well as removing impulses from the response. The following lemmas are introductory materials for the theorems developed later in this paper.

Lemma5: A control system is robustly stable, if and only if, the complementary sensitivity function fulfills the following inequality: [12]

$$\sup |\eta(s) \bar{l}_m| < 1 \quad (6)$$

Remark4: For IMC structure the complementary sensitivity function and the uncertainty can be computed as follows:

$$\eta(s) = \frac{qp}{1+q(p-\tilde{p})} = \frac{qp}{1+qp(1-1/D)} \quad (7)$$

$$\hat{l}_m(s) = D - 1 \quad (8)$$

Therefore, in this case condition (6) cannot be satisfied. Thus we need to modify the IMC structure or algorithm in order to gain a more tractable uncertainty profile. In the following section the SIMC filter is introduced and the proposed method is studied.

THE SIMC FILTER

The idea of augmenting the IMC compensator by an IMC filter can be extended to singular systems in a different manner. According to the previous discussions one way to overcome the obstacles in IMC of singular systems is to augment the compensator by an additional IMC filter, we call it SIMC. This filter have the same structure as the conventional IMC filter for step reference signals, and therefore, the IMC problem of singular systems consists of find-

ing two time constants; One for the conventional IMC which adjusts the closed loop performance, robustness and noise amplification; and one for the feasibility of robust control and impulse elimination of the singular plant. This is expectable for a singular system to require more parameters to be controlled, because a singular system is a general form of linear systems and cannot be treated by the same existing methods in standard systems. One advantage of SIMC is to solve the problem by introducing an additional filter without any need of complicated design procedures. Define the SIMC filter as a low pass filter:

$$f_2(s) = \frac{1}{(\tau s + 1)^m} \tag{9}$$

Lemma6: Define SIMC filter as stated in (9), therefore the closed loop system is strictly proper, if and only if:

$$m > -\sigma \tag{10}$$

In which, parameter σ denotes the relative degree of the plant.

Proof: Using (9-10) as the SIMC filter, the relative degree of plant/compensator becomes strictly proper. Therefore, by means of lemma2, the closed loop system is strictly proper. ■

Remark5: There is no need to introduce pole zero cancellation issues because SIMC filter cancels minimum phase zeros of the plant at most.

Lemma7: Together with SIMC filter the singular plant is capable of being robustly controlled, if (6) can be satisfied.

Proof: The new uncertainty profile have the following shape:

$$\hat{l}_m(s) = \frac{pf_2 - \frac{p}{D}}{\frac{p}{D}} = Df_2 - 1 \tag{11}$$

Now it is easy to choose SIMC filter such that the uncertainty profile is bounded.

■ Remark 6: Note that the real uncertainty profile between actual plant and assumed singular model is unchanged. SIMC manipulates only the mismatch between singular model and the implemented parallel strictly proper model of IMC. Also note that \hat{l}_m represents the uncertainty caused by singular system while l_m is the actual uncertainty.

Lemma8: The closed loop system with SIMC structure characterized by equations (9-11) is robustly stable, if and only if:

$$|f_1| < \frac{1}{|\tilde{p}\tilde{q}\hat{l}_m|} \tag{12}$$

Proof: The complementary sensitivity function can be stated as:

$$\eta(s) = \frac{qp}{1 + qf_2(p - p)} = qp = q\tilde{p}D = \tilde{q}f_1\tilde{p}Df_2 = \tilde{q}\tilde{p}f_1$$

Therefore condition (6) can be states as (12).

■ Remark7: The above lemma states an essential character of SIMC, the SIMC filter caused the uncertainty to remain in a disk shaped region and the robust stability criterion is then applicable to the problem. If one studies condition (6) with and without SIMC filter, it can be seen that SIMC filter imposes a bound on the uncertainty. Also choosing D as the inverse of SIMC filter, the uncertainty profile remains unchanged and the uncertainty caused by singular system will be zero as can be seen from (11).

When there are some model inaccuracies or disturbances, condition (12) cannot be met easily because a specific performance index is expected in the control objectives. In these situations a natural compromise exists and the victimization of performance is inevitable. Note that one can set the IMC filter to zero in order to satisfy (12) but this means open loop control of the system and therefore losing performance. The uncertainty bound generally increases with frequency. A natural routine for making the controller robust is to design a nominal H2-optimal controller according to performance specifications and then increasing the filter time constant to meet the desired robustness properties.

Theorem1: Assuming $D = f_2^{-1}$ then there is an IMC filter such that the closed loop system is robustly stable, and furthermore, the system exhibits robust performance at the zero frequency, if and only if:

$$\tilde{l}_m(0) < 1 \tag{13}$$

Proof: The IMC filter should satisfy (12) for robust stability, because of the structure selected for IMC filter, the maximum value for the filter is unity and it occurs at zero frequency. Therefore, for the nominal plant (12) can be satisfied only if the uncertainty upper bound is smaller than unity, and therefore, the necessary condition for the IMC filter to exist is (13). The sufficiency is obvious. ■

Remark8: Note that theorem 1 is an extension of the existing result in standard systems. Although the SIMC filter does not appear explicitly in the theorem, it has an essential role in the derivation of the theorem as well as lemmas. In other words, introducing the SIMC makes it possible to apply the existing frame for robust control to the singular systems.

Remark9: Theorem1 just considers the solvability of (12). In other words it studies the existence of an appropriate IMC filter which solves the robust control problem. In order to find such an IMC filter one should increase the time constant and check the robust stability criterion until it is satisfied.

Remark10: It should be noticed that there exists no constraint on the SIMC filter time constant and any positive time constant can be chosen. However when smoothness of the response is also a requirement, large time constant for SIMC filter is required, and when a fast response is desired, it is better to choose the time constant as small as possible. Note that if the SIMC filter time constant is larger than that of IMC filter and the plant dominant time constant, it will determine the closed loop time constant. In fact the closed loop time constant is the largest time constant among the plant, IMC filter and SIMC filter time constants. Because of robustness considerations SIMC filter time constant may be smaller than IMC filter time constant, and does not restrict the closed loop performance. It is not possible to decrease SIMC filter time constant as much as desired, since input noises may be amplified.

Remark11: Note that (13) means that steady state gains for the plant and model should be of the same sign. A little mismatch between plant and model steady state gain may cause instability if their sign were different. This a common drawback of robust control systems for plants with zeros near the origin. By a slight change of the zero location the closed loop may become unstable if the zero is near the origin.

Lemma9: Irregularity of closed loop occurs, if and only if:

$$cp = -1 \text{ for all } s$$

Proof: From the definition of regularity, a singular system is irregular if and only if:

$$|sE - A| \equiv 0 \quad (14)$$

In the frequency domain context of output feedback control systems the above determinant is the characteristic polynomial of system or the denominator of complementary sensitivity function. Write the closed loop transfer function as:

$$\eta(s) = \frac{cp}{1+cp} = \frac{\frac{N}{M}}{1+\frac{N}{M}} = \frac{N}{N+M}$$

According to (14) and (15) the closed loop system is irregular, if and only if:

$$N = -M$$

This can be rewritten as:

$$cp = -1 \text{ for all } s \quad (15)$$

The last equality also means an unsolvable algebraic loop in the simulation. ■

Corollary2: For a strictly proper plant/compensator, (15) does not occur because:

$$cp(s) = a_1s^{-1} + a_2s^{-2} + \dots \quad (16)$$

As a result for a strictly proper compensator/plant combination the regularity issue is not of concern. This corollary depicts the fact that why the regularity control objective is introduced only for singular systems and not for standard strictly proper ones.

In the following theorem we may introduce the interesting characteristics of the proposed algorithm.

Theorem2: The closed loop system with an appropriate IMC filter designed according to (12) is robustly strictly proper and robustly regular against all uncertainties described by (4).

Proof: Note that from theorem1, the closed loop robust stability and zero frequency performance are assured. The family of plants described by (4) all have a singularity index smaller than or equal to that of nominal plant. This can be shown as follows: assume that there is a plant in the family (4) that has a larger singularity index than the nominal plant. Then uncertainty profile can be written as:

$$\bar{l}_m = \frac{p - \tilde{p}}{\tilde{p}}$$

From the above assumption uncertainty will increase by frequency because it has an improper transfer function. Therefore, (4) cannot be satisfied as the uncertainty is unbounded. Moreover, for any plant being in family (4) the relative degree of SIMC filter is greater than or equal to the plant singularity index and thus the closed loop system is robustly strictly proper according to lemma2. Also note that

regularity of the plant is guaranteed by lemma8 because of strict properness of plant/compensator combination.

The following design procedure can be followed for robust internal model control of a singular plant.

Design Procedure:

1. Choose the polynomial D and set f_2 as its inverse. The polynomial time constant should be smaller than the dominant time constant of plant. According to nominal singular plant choose m such that the strictly properness of closed loop is guaranteed.
2. For the nominal plant check the feasibility of robust control having uncertainty profile as (4) according to (12), if satisfies design IMC filter for a good performance in nominal case.
3. Redesign SIMC filter for having better performance if required.

SIMULATION RESULTS

Simulating an improper system is not possible with the existing numerical methods, since simulation needs future data for computing the present state vector. This is why many Papers in the field of singular systems do not include any simulation examples or just simulate causal singular systems. However, if the closed loop system is proper, any simulating software can easily implement the closed loop system regardless of the inner unsolvable loops, which form singular systems in the inner parts of the closed loop system. In this paper, some illustrative simple examples are chosen in order to show the effectiveness of the proposed algorithm.

Example1: Consider the nonlinear system described by the following equations:

$$\begin{aligned} \dot{x}_1 &= -6x_1 + 2x_2 - u + x_1^2 \\ x_2 - u &= 0 \\ y &= x_1 + 2x_2 \end{aligned} \quad (17)$$

The output equation is the simplest form of output equation. Nonlinear output may occur in a singular system; however they are essentially treated in a similar manner. The algebraic part of a singular system denotes its limitations for having arbitrary initial conditions. The system described by (17) can be modeled by a standard state space system, too. For a nominal input of $u=9$, the equilibrium point will be:

$$\dot{x}_1 = \dot{x}_2 = 0 \Rightarrow x_1^* = 3, x_2^* = 9 \quad (18)$$

In case of nonlinear term in (17), the nonlinearity can be considered as an uncertainty, and not included in the linear model representation.

The process model may be considered as a bi-proper transfer function.

$$\tilde{p}(s) = \frac{2s+13}{(s+6)(0.1s+1)} \quad p(s) = \frac{2s+13}{(s+6)}$$

The compensator, IMC and SIMC filters may be chosen as:

$$f_2(s) = \frac{1}{(\tau s + 1)} = \frac{1}{D}$$

$$\tilde{q}(s) = \frac{(0.1s + 1)(s + 6)}{2s + 13}$$

$$f_1(s) = \frac{1}{(0.5s + 1)}$$

Closed loop system responses to different inputs with different initial conditions are shown in figures 1 to 3. Before change in the set point, initial condition is vanished and then the set point signal is tracked without any off set. The closed loop system is stable as the phase portrait shows and as it is strictly proper, a smooth response is attained. Closed loop system is able to follow any piecewise constant reference signal and reject constant disturbances. The main advantage of the presented control structure is to make closed loop, impulse free, even strictly proper, in a robust manner, i.e. regardless of uncertainty in the plant model.

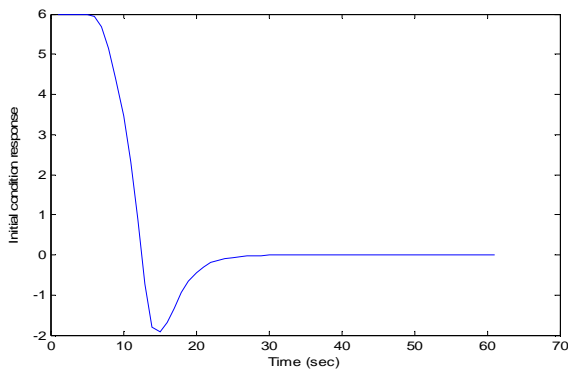


Figure 1: System response to initial condition

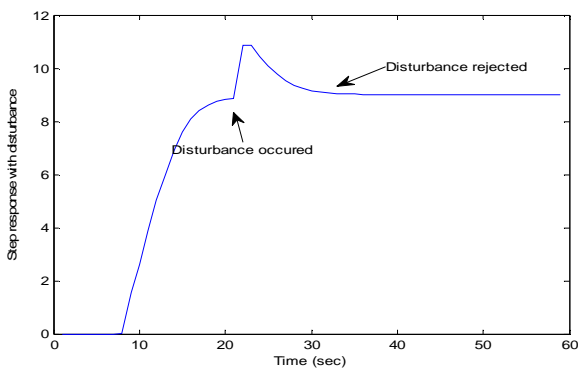


Figure 2: Set point and disturbance response

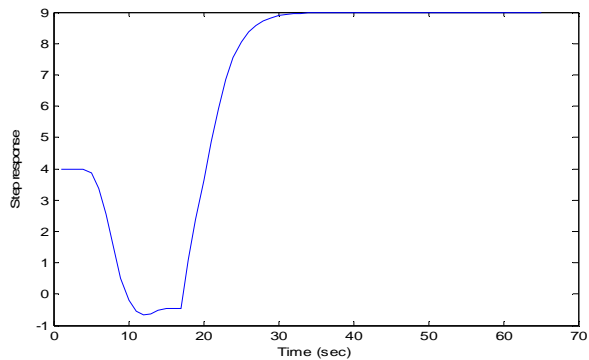


Figure 3: Set point response with initial condition

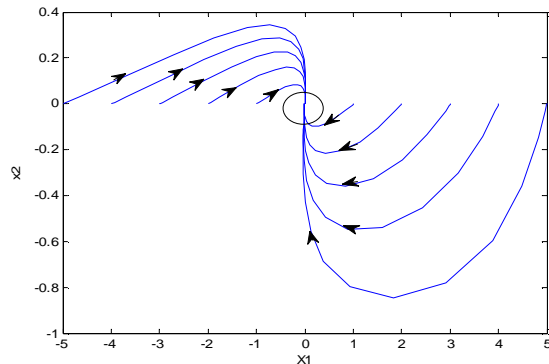


Figure 4: phase portrait of closed loop near the origin

This example shows that the closed loop system is strictly proper regardless of any model uncertainties while bounded and also it provides an example of robust stability and zero frequency performance design.

Example2: Consider a group of linear singular systems as described by the following set of transfer function.

$$p(s) = s^2 + \alpha s + 1$$

$$p_1(s) = s^2 + 3s + 1$$

$$p_2(s) = s^2 + 4s + 1$$

$$p_3(s) = s^2 + 1$$

The nominal plant (model) is assumed to be:

$$\tilde{p}(s) = s^2 + 2s + 1$$

The parallel model and compensator are:

$$\tilde{p}(s) = \frac{s^2 + 2s + 1}{(0.1s + 1)^3}$$

$$q(s) = f_1(s) \times \frac{(0.1s + 1)^3}{s^2 + 2s + 1}$$

It is aimed to design a single robust controller for all plants described above. The uncertainty norm is bounded for all of the models described in (19), however, its infinity norm is near the unity for case p4. Following figures depict closed

loop behavior in tracking step set point. Set point tracking is almost perfect even in presence of uncertainties. This characteristic also exists in conventional unit feedback control, e.g. PID controllers. However state space methods like [5] don't include this feature. In contrast to the aggressive nature of singular systems, closed loop response is smooth enough to ensure preventing any damage to the instruments.

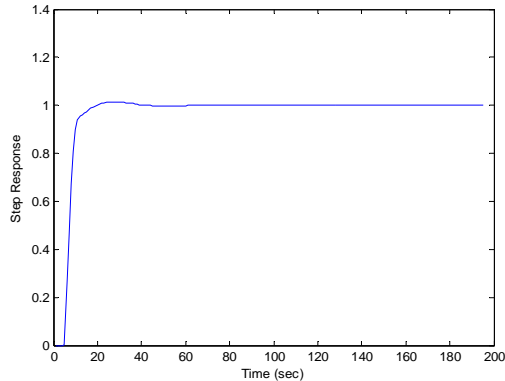


Figure 5: step response for p_1

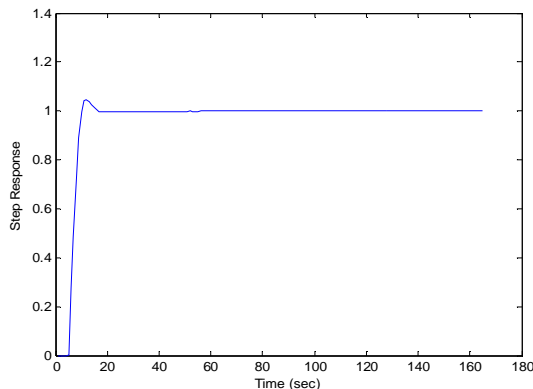


Figure 6: Step response for p_2

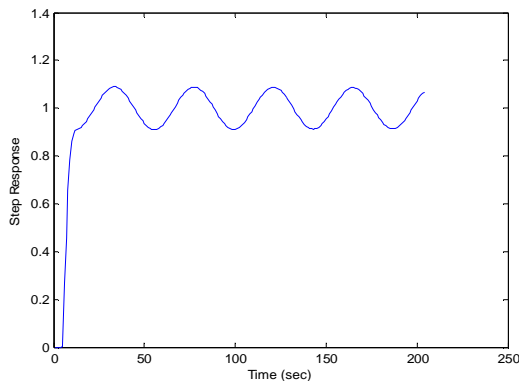


Figure 7: Step response for p_3

In the last case the oscillating behavior of plant is not included in the model and therefore closed loop response is not satisfactory. Note that while steady state gains of plant and model have the same sign, the closed loop is robustly stable. And while the uncertainty is bounded it is robustly strictly proper.

CONCLUSION

In this paper a new, effective and simple control scheme is proposed for robust internal model control of singular linear systems. The method has many advantages over the existing, state space methods including robust strict properness of the closed loop, avoiding algebraic loops, robust tracking of specific signals and the ability to robustly stabilize a larger group of singular systems comparing with other methods. Two simulation examples are included to depict the algorithm performance.

References

- [1] Mertzios, B.G. and F. Lewis, "Fundamental matrix of discrete singular systems", *Journal of circuits, systems and signal processing*, vol.8, NO.3, 1989.
- [2] Dai, L. "Singular control systems", Springer verlag, 1989.
- [3] Campbell, S.L, R. Nikoukhah and B.C. Levy, "Kalman filtering for general discrete time linear systems", *IEEE transactions on automatic control*, vol.44, NO.10, pp. 1829-1839,1999.
- [4] Luenberger, D. "Dynamic equations in descriptor form", *IEEE transactions on automatic control*, vol.AC.22, NO.3, June, 1977
- [5] Syrmos, V.L. and F.L. Lewis, "Robust eigen value assignment in generalized systems" *Proceedings of the 30th conference on decision and control*, England, 1991.
- [6] Fang, C.H., L.Lee and F.R. Chang, "Robust control analysis and design for discrete time singular systems", *Automatica*, Vol.30, NO.11, pp.1741-1750, 1994.
- [7] Xu, S. , C. Yang, Y. Neu and J. Lam, "Robust stabilization for uncertain discrete singular systems", *Automatica*, Vol.37, pp.769-774, 2001.
- [8] Mukundan, R. and W. Dayawansa, "Feedback control of singular systems- proportional and derivative feedback of the state", *International journal of systems science*, vol.14, NO.6, pp.615-632, 1983.
- [9] Chu, D.L., H.C. Chan and D.W. Ho, " Regularization of singular systems by derivative and proportional output feedback", *Siam journal of matrix nalysis and applications*, vol.19, NO.1, pp.21-38, January, 1998.
- [10] Xu, S. and J. Lam, " Robust stability and stabilization of discrete singular systems: An equivalent characterization", *IEEE transaction on automatic control*, vol.49, NO.4, April 2004.
- [11] Hou, M. "Controllability and elimination of impulsive modes in descriptor systems", *IEEE transactions on automatic control*, vol.49, NO.10, 2004.
- [12] Morari, M. and E. Zafiriou, "Robust process control", Prentice Hall, 1989.

An Integrated Partial Backlogging Inventory Model having Weibull Demand and Variable Deterioration rate with the Effect of Trade Credit

P.K. Tripathy, S. Pradhan

Abstract - Demand considered in most of the classical inventory models is constant, while in most of the practical cases the demand changes with time. In this article, an inventory model is developed with time dependent two parameter weibull demand rate whose deterioration rate increases with time. Each cycle has shortages, which have been partially backlogged to suit present day competition in the market. Also the effect of permissible delay is also incorporated in this study. The total cost consists of ordering cost, inventory holding cost, shortage / backordering cost, lost sale cost and deterioration cost are formulated as an optimal control problem using trade credit policy. Optimal solution for the model is derived and the trade credit on the optimal replenishment policy are studied with the help of numerical examples.

Index Terms- Inventory, Shortages, partial backlogging, weibull demand, trade credit , variable deterioration, replenishment

1. INTRODUCTION

In the classical inventory economic order quantity (EOQ) model, it was tacitly assumed that the customer must pay for the items as soon as the items are received. However, in practices or when the economy turns sour, the supplier frequently offers its customers a permissible delay in payments to attract new customer. In today's business transaction, it is frequently observed that a customer is allowed some grace period before settling the accounts with the supplier or the producer. The customer does not have to pay any interest during this fixed period but if the payment gets delayed beyond the period interest will be charged by the supplier. This arrangement comes out to be very advantageous to the customer as he may delay the payment till the end of the permissible delay period. During the period, he may sell the goods, accumulate revenues on the sales and earn interest on that revenue. Thus it makes economic sense for the customer to delay the payment of the replenishment account upto the last day of the settlement period allowed by the supplier on the producer. Similarly for supplier, it helps to attract new customer as it can be considered some sort of loan. Furthermore, it helps in the bulk sale of goods and the existence of credit period serves to reduce the cost of holding stock to the user, because it reduces the amount of capital invested in stock for the duration of the credit period. So, the concept of permissible delay in payments beneficial both for supplier and customer.

In some real life situation there is a part of demand which can not be satisfied from the current inventory, leaving the system in stock out. In these systems two situations are mainly considered, all customers wait until the arrival of the next order (Complete back order case) or all customers leave the system (lost sales case). However, in practical, some customers are able to wait for the next order to satisfy their demands during the stock out period, while others do not wish to or can not wait and they have to fill their demands from other sources. This situation is modeled by consideration of partial back- ordering in the formulation of the mathematical model. Wee (1999) develops a deterministic inventory model with quantity discount, pricing and partial backlogging when the product in stock deteriorates with time according a weibull distribution. Teng (2002) presents on EOQ model under the conditions of permissible delay. Chen et al. (2003) establish an inventory model having weibull deterioration and time varying demand. Wu et al. (2003) considered an inventory model where deteriorating rate and demand rate are follows the Chen's model (2003) where shortages are permitted. Papachristos and Skouri (2003) present a production inventory model with production rate, product demand rate and deteriorating rate, all considered as functions of the time. Their model follows shortages and the partial rate is a hyperbolic function of the time up to the order point. They propose an algorithm for finding the solution of the problem. Abad (2003) considers the problem of determining the optimal price and lot size for reseller in which unsatisfied demand is partially backordered. There are several interesting papers related to partial backlogging and trade credits viz. park (1982), Jamal et al. (1997) , Lin et al. (2000), Dye et al. (2007) and their references.

With this motivation, in this paper an attempt is made to formulate an inventory model in time dependence of demand follows a two parameters weibull type with time variable deterioration where unsatisfied

1. P.G. Department of Statistics, Utkal University, Bhubaneswar-751004, India, msccompssc@gmail.com
2. Dept. of Mathematics, Orissa Engineering College, Bhubaneswar-751007, India, suchipradhan@yahoo.com

demand is partially backlogged and delay payment is allowed. With the help of this model, supplier can easily extract its order quantity, cycle period for shortages as well as payment time to reduce the total inventory cost

(Fig-1)

2. NOTATION AND ASSUMPTIONS:

To develop the proposed model, the following notations and assumptions are used in this article.

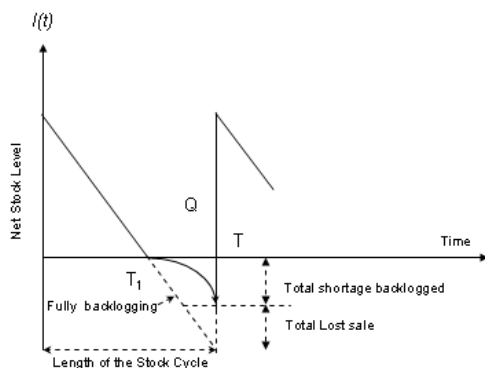
2.1 Notations

- $I(t)$: Inventory level at time t
- $Q(t)$: Order quantity at time $t = 0$
- A, h, C_1, C_2, C_3 : Ordering cost, holding cost, deteriorating cost, shortage cost for backlogged items and the unit cost of lost sales respectively. All of the cost parameters are positive constants.
- I_e : Interest which can be earned
- I_c : Interest charges which invested in inventory, $I_c \geq I_e$
- M : Permissible delay in setting the amounts, $0 < M < T$
- T : Length of the replenishment cycle
- T_1 : Time when inventory level comes down to zero, $0 < T_1 < T$
- TIC_1 : Total inventory cost when $M < T_1$
- TIC_2 : Total inventory cost when $T_1 < M < T$

2.2 Assumptions:

1. The inventory system involves only one item
2. Replenishment occurs instantaneously on ordering i.e. lead time is zero and it takes place at an infinite rate.
3. The demand rate of any time t is $\alpha \beta t^{\beta-1}$ two parameter weibull type, where $0 < \alpha \ll 1, \beta > 0$ are called scale and shape parameter respectively.
4. $Q(t) = vt$ is the variable deterioration rate, where $0 < v \ll 1$.
5. Shortages are allowed and unsatisfied demand is backlogged at the rate of $\frac{1}{1 + \delta(T-t)}$

The backlogging parameter δ is a positive constant, where $T_1 \leq t < T$.



3. MODEL FORMULATION:

In this model two parameter weibull type of demand is considered with variable rate of deterioration. Depletion of the inventory occurs due to demand (supply) as well as due to inventory i.e. during the period (0,T) and shortages occurs during period (T₁, T). The differential equation of inventory level I (t) w.r.t. time is given by

$$\frac{dI(t)}{dt} + vtI(t) = -\alpha \beta t^{\beta-1} \tag{1}$$

With boundary condition $I(0) = Q$ and $I(T_1) = 0$

The solution of equation (1) is

$$I(t) = \frac{\alpha v}{\beta + 2} (t^{\beta+2} - T_1^{\beta+2}) + \tag{2}$$

$$\alpha (T_1^\beta - t^\beta) - \frac{\alpha vt^2 T_1^\beta}{2}, 0 < t < T$$

$$\text{and } Q = \alpha T_1^\beta - \frac{\alpha v \beta T_1^{\beta+2}}{\beta + 2} \tag{3}$$

Similarly the differential equation for shortages is given by

$$\frac{dI(t)}{dt} = -\frac{\alpha \beta t^{\beta-1}}{1 + \delta(T-t)} \tag{4}$$

The solution of equation (4) for shortage is

$$I(t) = \alpha t^\beta (\delta T - 1) - \frac{\alpha \beta \delta t^{\beta+1}}{\beta + 1} \tag{5}$$

The deteriorating cost during one cycle is

$$DC = C_1 \left[Q - \int_0^{T_1} \alpha \beta t^{\beta-1} dt \right] = \frac{-C_1 \alpha v T_1^{\beta+2}}{\beta + 2} \tag{6}$$

The holding cost for carrying inventory over the period (0,T) is given by

$$HC = \int_0^{T_1} h I(t) dt = h \left[\frac{\alpha \beta T_1^{\beta+1}}{\beta + 1} - \frac{\alpha v T_1^{\beta+3} (\beta + 9)}{6(\beta + 3)} \right] \tag{7}$$

The total amount of shortage cost during the period (T₁, T) is given by

$$SC = -C_2 \int_{T_1}^T I(t) dt = C_2 \left[\frac{\alpha (\delta T - 1)}{\beta + 1} (T_1^{\beta+1} - T^{\beta+1}) + \frac{\alpha \beta \delta}{(\beta + 1)(\beta + 2)} (T^{\beta+2} - T_1^{\beta+2}) \right] \tag{8}$$

The amount of lost cost during the period (0,T) is given by

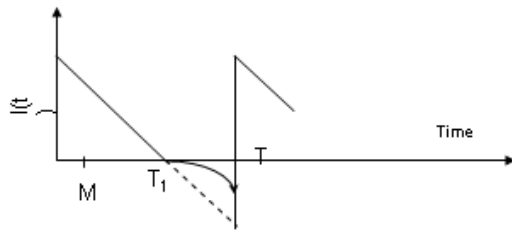
$$LC = C_3 \int_{T_1}^T \left[1 - \frac{\alpha \beta t^{\beta-1}}{1 + \delta(T-t)} \right] dt = C_3 \left[\frac{\delta \alpha T^{\beta+1}}{\beta + 1} - \delta \alpha T_1^\beta \left(T - \frac{\beta T_1}{\beta + 1} \right) \right] \tag{9}$$

Case – I $0 < M < T_1$

In this situation since the length of period with positive stock is larger than the credit period, the buyer can use the sale revenue to earn interest at an annual rate I_e in $(0, T_1)$

The interest earned IE_1 , is

$$IE_1 = PI_e \int_0^M \alpha \beta t^\beta dt = \frac{PI_e \alpha \beta M^{\beta+1}}{(\beta+1)} \tag{10}$$



(Fig-2)

However, beyond the credit period, the unsold stock is supposed to be financed with an annual rate I_c and the interest charged IC_1 , is

$$IC_1 = PI_c \int_M^{T_1} I(t) dt$$

$$= PI_c \left[\frac{\alpha \beta T_1^{\beta+1}}{\beta+1} - \frac{\alpha v (\beta+9) T_1^{\beta+3}}{6(\beta+3)} - \frac{\alpha v}{\beta+2} \left(\frac{M^{\beta+3}}{\beta+3} - M T_1^{\beta+2} \right) \right]$$

$$+ \alpha \left[T_1^\beta M - \frac{M^{\beta+1}}{\beta+1} - \frac{\alpha v T_1^\beta M^3}{6} \right] \tag{11}$$

Total inventory cost per unit time is given by

$$TIC_1 = \frac{1}{T} [A + HC + SC + LC + DC + IC_1 - IE_1]$$

$$= \frac{1}{T} \left[A + h \left\{ \frac{\alpha \beta T_1^{\beta+1}}{\beta+1} - \frac{\alpha v (\beta+9) T_1^{\beta+3}}{6(\beta+3)} \right\} + C_2 \left\{ \frac{\alpha (\delta T - 1)}{(\beta+1)} (T_1^{\beta+1} - T^{\beta+1}) \right. \right.$$

$$+ \left. \frac{\delta \alpha \beta (T^{\beta+2} - T_1^{\beta+2})}{(\beta+1)(\beta+2)} \right\} + C_3 \left\{ \frac{\delta \alpha T^{\beta+1}}{\beta+1} - \delta \alpha T_1^\beta \left(T - \frac{\beta T_1}{\beta+1} \right) \right\}$$

$$- C_1 \left\{ \frac{\alpha v T_1^{\beta+2}}{\beta+2} \right\} + PI_c \left\{ \frac{\alpha \beta T_1^{\beta+1}}{\beta+1} - \frac{\alpha v (\beta+9) T_1^{\beta+3}}{6(\beta+3)} \right.$$

$$- \left. \frac{\alpha v}{\beta+2} \left(\frac{M^{\beta+3}}{\beta+3} - M T_1^{\beta+2} \right) - \alpha \left(T_1^\beta M - \frac{M^{\beta+1}}{\beta+1} \right) + \frac{\alpha v M^3 T_1^\beta}{6} \right\}$$

$$- \left. \frac{PI_e \alpha \beta M^{\beta+1}}{\beta+1} \right] \tag{12}$$

The solutions for the optimal values of T_1 and T can be found by solving the following equations simultaneously.

$$\frac{\partial TIC_1}{\partial T} = 0 \text{ and } \frac{\partial TIC_1}{\partial T_1} = 0 \tag{13}$$

Provided they satisfy the conditions

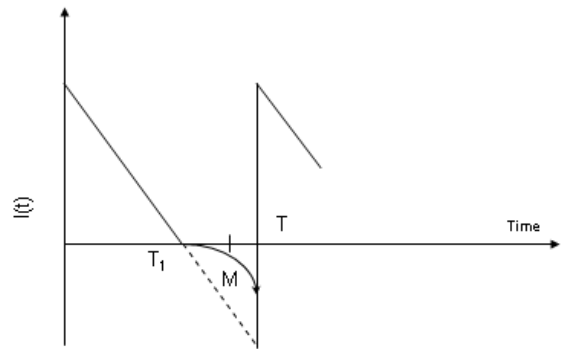
$$-\frac{\partial^2 TIC_1}{\partial T^2} > 0 \text{ and } \frac{\partial^2 TIC_1}{\partial T_1^2} > 0 \tag{14}$$

Case – II: $T > M > T_1$

In this situation interest charged $IC_2=0$ and the interest earned per time unit is

$$IE_2 = PI_e \left[\int_0^T \alpha \beta t^\beta dt + \alpha \beta T^\beta (M - T) \right]$$

$$= PI_e \left[\alpha \beta M T^\beta - \frac{\alpha \beta^2 T^{\beta+1}}{\beta+1} \right] \tag{15}$$



(Fig-3)

Then the total inventory cost per unit time is given by

$$TIC_2 = \frac{1}{T} [A + HC + SC + LC + DC + IC_2 - IE_2]$$

$$= \frac{1}{T} \left[A + h \left\{ \frac{\alpha \beta T_1^{\beta+1}}{\beta+1} - \frac{\alpha v (\beta+9) T_1^{\beta+3}}{6(\beta+3)} \right\} \right.$$

$$+ C_2 \left\{ \frac{\alpha (\delta T - 1)}{\beta+1} (T_1^{\beta+1} - T^{\beta+1}) + \frac{\delta \alpha \beta (T^{\beta+2} - T_1^{\beta+2})}{(\beta+1)(\beta+2)} \right\}$$

$$+ C_3 \left\{ \frac{\delta \alpha T^{\beta+1}}{\beta+1} - \delta \alpha T_1^\beta \left(T - \frac{\beta T_1}{\beta+1} \right) - C_1 \left\{ \frac{\alpha v T_1^{\beta+2}}{\beta+2} \right\} \right.$$

$$- \left. PI_e \left[\alpha \beta M T^\beta - \frac{\alpha \beta^2 T^{\beta+1}}{\beta+1} \right] \right] \tag{16}$$

The solutions for the optimal values of T_1 and T can be found by solving the following equations simultaneously.

$$\frac{\partial^2 TIC_2}{\partial T^2} = 0 \text{ and } \frac{\partial^2 TIC_2}{\partial T_1^2} = 0 \tag{17}$$

Provided they satisfy the conditions

$$\frac{\partial^2 TIC_2}{\partial T^2} > 0 \text{ and } \frac{\partial^2 TIC_2}{\partial T_1^2} > 0 \tag{18}$$

Numerical Examples:

Case I : A=50, h=2, c₁=2, C₂=0.8, C₃=2.0, α =0.0002, v=0.001, δ =0.5, p=0.5, l_c= 0.2, l_e=0.1, m =0.05

β	T ₁	T	Q	TIC ₁
0.8	1.90535191	106.840667	0.000335	0.718680
1.0	2.2908332	77.122890	0.000459	1.130076
1.5	2.917128	40.368203	0.001	2.662553
2.0	3.01722969	24.922125	0.001829	4.927522
2.5	2.77056753	17.154589	0.002566	7.866730
3.0	2.4129991	12.736923	0.00282	11.374538
3.5	2.07280312	9.991002	0.002571	15.332234
4.0	1.78842979	8.166399	0.00205	19.628761
4.5	1.56061254	6.889604	0.001485	24.170034
5.0	1.379736015	5.958621	0.001001	28.880925

Case II :

A=50, h=2, c₁=2, C₂=0.8, C₃=2.0, α =0.0002, v=0.001, δ =0.5, p=0.5, l_e=0.1, m =0.5

β	T ₁	T	Q	TIC ₂
0.8	0.4998753	1.257346	0.000115	39.766547
1.0	0.41972384	1.066637	0.000084	46.876526
1.5	0.398675	1.051521	0.00005	46.962943
2.0	0.37985845	1.047747	0.000029	47.721766
2.5	0.359786885	1.007243	0.000016	49.640806
3.0	0.34985674	0.990081	0.000009	50.501263
3.5	0.29985764	0.85216	0.000003	58.674548
4.0	0.2789685	0.797059	0.000001	62.730769
4.5	0.2756759	0.791841	0.000001	63.144167
5.0	0.2679865	0.772816	0.0000	64.698610

4. CONCLUDING REMARKS:

In this paper, an EOQ inventory model is considered for determining the optimal replenishment schedule under variable deterioration with two parameter weibull type demand where shortages are partially back logged. Also the proposed model in-cooperates other realistic phenomenon and practical features such as trade credit period. The credit policy in payment has become a very powerful tool to attract new customers and a good incentive policy for the buyers. In keeping with this reality, these factors are incorporated into the present model. The model is very useful in retail business. In real market situations demand rate may be increased, decreased or constant. Keeping these consideration, we have assumed here that the time dependence of demand follows the two parameters weibull distributions where depending upon the values of the parameter β (<1, >1 and >2), the degree of non-linearity varies. (see Appendix I) Numerical examples are presented to justify the claim of each case

of the model analysis by obtaining the optimal inventory length ,optimal economics order quantity, shortage time period and also calculated the total variable cost. From the numerical examples it is found that for high values of the parameter β, zero terminal inventory level provides higher unit of the total variable cost. Thus for simplicity regarding the applicability of the model it is not preferable to adopt zero-terminal inventory level policy.

The proposed model can be extended in numerous ways. For example we may consider the effect of inflation and time value of money by taking discount rate.

APPENDIX - I

The demand rate function D(t) =αβt^{β-1}, α, β > 0 follows two parameters weibull distribution in time t.

Thus functional form represents increasing, decreasing and constant demand for different values of the parameter β

We have

$$\frac{dD(t)}{dt} = \alpha\beta(\beta - 1)t^{\beta-2} \tag{19}$$

and
$$\frac{d^2 D(t)}{dt^2} = \alpha\beta(\beta - 1)(\beta - 2)t^{\beta-3} \tag{20}$$

- i) for 0<β<1, $\frac{D(t)}{dt} < 0$ and $\frac{d^2 D(t)}{dt^2} > 0$ the demand rate decreased with time at an increasing rate.
- ii) β=1, the demand rate becomes steady over time.
- iii) 1< β ≤ 2, $\frac{d D(t)}{dt} > 0$, $\frac{d^2 D(t)}{dt^2} \leq 0$, then the demand increases with time at decreasing rate.

Case (i) is applicable to the seasonal products towards the end of the seasons.

Case (ii) is applied to the consumer goods long established in the market by their brand name.

Case (iii) is appropriate for some fashionable product lunched in the market

If β >2 then $\frac{d D(t)}{dt} > 0$ and $\frac{d^2 D(t)}{dt^2} > 0$, which implies that the demand will go on increasing with time at an increasing rate. Thus there will be an accelerated growth in demand, which is not very realistic.

REFERENCE

[1] A.M.Jamal,B.R.Sarker and S.Wang (1997).An ordering policy for deteriorating items with allowable shortage and permissible delay in payment, Journal of the Operational Research Society,48,826-833.

- [2] Jimm-Tsair Teng and Liang-yuh ouyang(2005).An EOQ model for Deteriorating items with Power-Form Stock-Dependent demand ,Information and Management Sciences, 16,1-16.
- [3] S.K.Goyal and B.C.Giri(2001).Recent trends in modeling of deteriorating inventory, European Journal of Operational Research,134,1-16.
- [4] R.B Covert and G.S.Philip(1973).An EOQ model with weibull distribution deterioration, AIIE Transaction,5,323-326.
- [5] T.K.Datta and K.Paul(2001).An inventory system with stock- dependent, price sensitive demand rate, Production Planning and Control, 12, 13-20.
- [6] I. T. Teng(2002). An Economic order quantity under conditions of permissible delay in payments, Journal of the Operational Research Society. 53, 915-918.
- [7] J. M. Chen and S. C. Lin(2003). Optimal replenishment scheduling for inventory items with weibull distributed deterioration and time varying demand, Journal of Information and Optimization Sciences, 24 ,1-21.
- [8] J. W. Wu and W. C. Lee(2003). An EOQ inventory model for items with weibull distributed deterioration, shortages and time –varying demand. Information and optimization Science 24, 103-122 .
- [9] S. Papachristos and K. Skouri(2003). An inventory model with deteriorating items, quantity discount , pricing and time –dependent partial backlogging, International Journal of Production Economics, 83, 247-256.
- [10] C. Y. Dye , L. Y. Ouyang and T. P. Hsieh(2007). Deterministic inventory model for deteriorating items with capacity constraint and time proportional backlogging rate, European Journal of Operational Research, 178, 789-807.
- [11] K.S Wu, L.Y ouyang and C.T Yang(2006).An optimal replenishment policy for non-instantaneous deteriorating items stock dependent demand and partial back logging , International Journal of Production Economics,101,369-384.
- [12] P.L.Abad(2003).Optimal pricing and lot-sizing under conditions of perishability, finite production and partial back ordering and lost sale, European Journal of Operational Research, 144, 677-685.
- [13] S. K. Park (1982) Inventory model with partial backorders. International Journal of Systems Science, 13, 1313-1317.
- [14] H. M. Wee (1999). Deteriorating inventory model with quantity discount, pricing and partial backordering. International Journal of Production Economics, 59, 511-518 .
- [15] C. Lin, B. Tan and W. C. Lee(2000). An EOQ model for deteriorating items with time varying demand and shortages, International Journal of systems Science, 31, 391-400.

Survey of Current and Future Trends in Security in Wireless Networks

Vikas Solomon Abel

Abstract— Security has always been a key issue with all the wireless networks since they have no physical boundaries. Many existing and evolving threats which must be considered to ensure the countermeasures are able to meet the security requirements of the environments for which it is expected to be deployed.

Index Terms— Bluetooth, Wi-Fi, WiMax, Wireless Sensors, WLAN

1 Wireless Technologies

Wireless Broadband is a fairly new technology that provides high-speed wireless internet and data network access over a wide area. The three most important broadband wireless technologies are IEEE 802.11, IEEE 802.16, and Wireless Mesh Network (WMN). In Wireless Broadband Networks Privacy and Integrity of data are key considerations for both broadband networks. Authentication and Security to prevent unwanted access to critical data or services are necessary for the effective operation of any broadband network. The motivation behind Wireless Broadband technology was the support of mobile clients within a certain range and to provide wireless network infrastructure in such places where wired networks are not feasible or possible.

WLAN which is IEEE 802.11 technology consist of two important components, i.e. nodes and an Access Point (AP). The AP provides wireless connectivity to nodes as well as function as a bridge between the nodes and wired network infrastructure. Probe request frames are used by the clients to discover a wireless network. If a wireless network exist then the AP respond with Probe response frame. The AP (Access Point) which provides the strongest signal is selected. If the AP is overloaded then the nodes connect with another nearby AP although the signal strength is relatively weak [1]. Typical applications of WLAN are campus and organizational networking; back haul for public safety and industrial control networks [2]. WLAN may be an Infrastructure WLAN where client devices communicate wirelessly to a wired Local Area Network (LAN) such as Ethernet, via Access Points or an Ad-Hoc WLAN in which devices or stations communicate directly with each other, without using Access Points or a

connection to the wired network. For wireless devices in a WLAN to communicate with each other, they must all be configured with the same SSID. Wireless devices have a default SSID that is set at the factory. Some wireless devices refer to the SSID as network name.

Wi-Fi or Wireless Fidelity is to connect to the internet at high speeds. Wi-Fi enabled devices operates in unlicensed spectrum. They use radio technologies within the range of an access point for data communication which are based on the IEEE 802.11 standards. The benefits of using Wi-Fi in the last mile are that, the client device is so cheap due to large volume of production [3].

Sensor networks often have one or more points of centralized control called base stations. A base station is typically a gateway to another network, a powerful data processing or storage center, or an access point for human interface. They can be used as a nexus to disseminate control information into the network or extract data from it. They may also be referred to as sinks. Base stations are typically more powerful than sensor nodes. Sensors are constrained to use lower-power, lower-bandwidth, shorter-range radios. Sensor nodes form a multihop wireless network to allow sensors to communicate to the nearest base station.

Aggregation points help in reducing the total number of messages sent thus saving energy, sensor readings from multiple nodes may be processed at one of many possible aggregation points. An aggregation point may collect sensor readings from surrounding nodes and forwards a single message representing an aggregate of the values. It is also possible that every node in the network functions as an aggregation point, delaying transmission of an outgoing message until a sufficient number of incoming messages

have been received and aggregated. Power management in sensor networks is vital [4].

WMN is an emerging broadband wireless technology which consists of mesh routers and mesh nodes. The mesh routers form the backbone and provide connectivity to mesh nodes. With some of the mesh routers in WMN the functionality of gateways is provided for internet connectivity to mesh nodes. The WMN nodes act both as a client and as a relay router for forwarding traffic for other nodes. In most of the cases, the WMN form partial mesh topology, which means that there are many routes to every node, so if any route is not working then the data traffic can be sent through other route [2].

Bluetooth is a universal radio interface in the 2.45 GHz frequency band that enables portable electronic devices to connect and communicate wirelessly via short-range, ad hoc networks. Each unit can simultaneously communicate with up to seven other units per piconet. Moreover, each unit can simultaneously belong to several piconets [5].

Bluetooth uses a radio technology called frequency-hopping spread spectrum. A Bluetooth device will use many different and randomly chosen frequencies every second to minimize the probability of other devices using the same frequency and to minimize interference time. For battery operated devices Bluetooth is ideal. It does not rely on the user input since it can automatically detect and communicate with other Bluetooth devices. The frequency of Bluetooth capable devices ranges from 2.402 GHz to as high as 2.480 GHz, a frequency range specifically reserved by international agreement for ISM or medical, industrial and scientific devices.

WiMAX, an acronym for Worldwide Interoperability for Microwave Access, is a telecommunications technology that provides fixed and fully mobile internet access. The IEEE 802.16 standard forms the basis of . Clarification of the formal names are as follow:

802.16-2004 is also known as 802.16d, which refers to the working party that has developed that standard. Since it has no support for mobility it is sometimes referred to as "Fixed WiMAX". 802.16e-2005, is also known as 802.16e, is an amendment to 802.16-2004. It introduced support for mobility, among other things and is therefore also known as "Mobile WiMAX". 802.16j-2009, the Multihop Relay specification for 802.16.

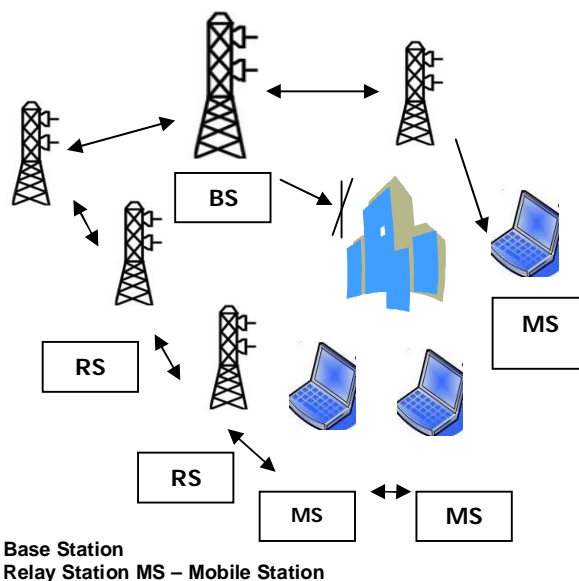


Fig 1. Multi-hop Relay Network architectures
Source : [23]

2 General Attacks on Wireless Networks

General, attacks on wireless networks may be divided into four basic categories: passive attacks, active attacks, man-in-the-middle attacks, and jamming attacks.

A passive attack occurs when someone listens to or eavesdrops on network traffic. A passive attack on a wireless network may not necessarily be malicious in nature. They are very difficult to detect.

In passive attacks the attacker is not concerned with the routing protocol. It only eavesdrops on the routing traffic and endeavours to get valuable information like node hierarchy and network topology from it. For example, if a route to a particular node is requested more than other nodes, the attacker might predict that the node is important for the operation of the network, and putting it out could bring down the entire network. Even when it is not possible to isolate the precise position of a node, one may be able to determine information about the network topology by analysing the contents of routing packets. This attack is very difficult to detect and prevent [6].

In active attacks, the aggressor node has to spend some of its energy in order to carry out the attack. Nodes that perform active attacks with the aim of disrupting other nodes by causing network outage are considered to be harmful, while nodes that perform passive attacks with the

aim of saving battery life for their own communications are considered to be selfish. The harmful or malicious nodes can change the routing information by modifying, fabricating or impersonating nodes and thereby disrupting the correct functioning of a routing protocol [6].

Once an attacker has gained enough information from the passive attack, the hacker can then launch an active attack against the network. There are a large number of active attacks that a hacker can launch against a wireless network. Some examples of the attacks are Spoofing attacks, Unauthorized access, Flooding attacks etc. Spoofing occurs when a malicious node misrepresents its identity by altering its MAC or IP address in order to mislead the information about the network topology. It can also create routing loops etc.

3 General Physical layer Attacks

3.1 Jamming Attacks on Wireless Networks

Jamming is a attack specific to wireless networks. Jamming occurs when spurious RF frequencies interfere with the operation of the wireless network. In some cases, the jamming may be unintentional and is caused by the presence of other devices such as cordless phones, that operate in the same frequency as the wireless network. Intentional and malicious jamming occurs when an attacker analyzes the spectrum being used by wireless networks and then transmits a powerful signal to interfere with communication on the frequencies discovered.

3.2 Scrambling

Scrambling is a type of jamming attack for short intervals to disorder targeted frames (mostly management messages) which ultimately lead to network failure.

3.3 Water Torture Attack

In this attack the attacker pushes a Subscriber Station (SS) to drain its battery or consume computing resources by sending bogus frames [7].

3.4 Man in the Middle Attack

In this attack an attacker intercepts the valid frames and then intentionally resends the frames to the target system. A wireless – specific variation of man-in-the-middle attack is placing a rogue access point within range of wireless stations. If the attacker knows the SSID in use by the network, he can gain information such as the key information, authentication requests and so on. It may involve spoofing an IP address, changing a MAC address to emulate another host, or some other type of modification.

3.5 Denial of Service Attack

In a "denial-of-service" attack an attacker may try to prevent an authentic user from using a service. This may be done by "flooding" a network, disrupt connections between two machines etc. The common method is to flood a network with degenerate or faulty packets and hence denying access to the legitimate traffic.

3.6 Physical Tampering

In this attack the physical device may be tampered and sensitive information can be extracted from it.

4 General Link layer Attacks

4.1 Unencrypted Management Communication

Almost all the IEEE 802.16 management messages are still sent unencrypted. The IEEE 802.16-2004 standard does not provide any capability to encrypt management messages.

4.2 Masquerading threat

By intercepting the management messages an attacker can use the hardware address of another registered device. Once this is successful an attacker can turn a Base Station (BS) into a Rogue Base Station [7]. A rouge WiMAX base station pretends to be a valid base station, and then drops/eliminates all/some of the packets.

4.3 Threat due to Initial Network Entry

In IEEE 802.16j-2009 no integrity protection for management messages can be made in case of multicast transmissions and in case of initial network entry by a new candidate node [8].

5 General Network layer Attacks

The main network-layer operations in MANETs are ad hoc routing and data packet forwarding, which interact with each other and fulfill the functionality of delivering packets from the source to the destination. Routing messages in Ad Hoc routing protocols exchange routing messages between nodes and maintain routing states at each node accordingly. Data packets are forwarded by intermediate nodes along an established route to the destination based on the routing states. Both routing and packet forwarding operations are vulnerable to malicious attacks, leading to various types of malfunction in the network layer. Network-layer vulnerabilities generally fall into one of two categories: routing attacks and packet forwarding attacks, based on the target operation of the attacks [9].

In the context of DSR [11], the attacker may modify the source route listed in the RREQ or RREP packets by deleting a node from the list, switching the order of nodes in the list, or appending a new node into the list [12]. In AODV [10] are used, the attacker may advertise a route with a smaller distance metric than its actual distance to the destination, or advertise routing updates with a large sequence number and invalidate all the routing updates from other nodes [13]. By attacking the routing protocols, the attackers can pull traffic toward certain destinations in the nodes under their control, and cause the packets to be forwarded along a route that is not the optimal route or which may even be nonexistent. The attackers can also create routing loops in the network [9].

Attacks may be launched in addition to routing attacks such as in packet forwarding operations. These attacks do not disrupt the routing protocol but poison the routing states at each node. Instead, they cause the data packets to be delivered in a way that is intentionally inconsistent with the routing states. The attacker may drop or modify the contents of the packets or duplicate the packets it has already forwarded. Attacker may also inject large amount of bogus packets into the network which wastes a significant portion of the network resources and introduces severe wireless channel contention and network congestion in the MANET [9].

Routing protocols for ad-hoc networks are based on the assumption that intermediate nodes do not maliciously change the protocol fields of messages passed between nodes. This assumed trust permits malicious nodes to easily generate traffic subversion and denial of service (DoS) attacks. Attacks using modification are generally targeted against the integrity of routing computations and so by modifying routing information an attacker can cause network traffic to be dropped, redirected to a different destination, or take a longer route to the destination increasing communication delays. Forged routing packets may be sent to other nodes to divert traffic to the attacker or to some other node. The intention is to create a black hole by routing all packets to the attacker and then discarding it. As an extension to the black hole, an attacker could build a grey hole, in which it intentionally drops some packets but not others, for example, forwarding routing packets but not data packets. A more subtle type of modification attack is the creation of a tunnel (or wormhole) in the network between two colluding malicious nodes linked through a private network connection [6]. A brief description of the attacks is as follows -

5.1 Blackhole Attack

In networking, black holes refer to places in the network where incoming traffic is silently discarded (or "dropped"), hiding from the source the information that the data did not reach its intended recipient. The black holes themselves are invisible, and can only be detected by monitoring the lost traffic. An attacker creates forged packets to impersonate a valid mesh node and subsequently drop packets [14].

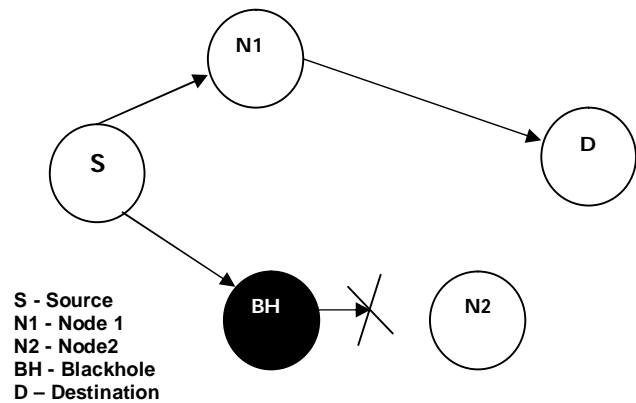


Fig 2. Blackhole Attack
Source : [22]

The properties of Blackhole attack is that firstly, the Blackhole node exploits the ad hoc routing protocol, such as AODV, to advertise itself as having a valid route to a destination node, even though the route is false, with the intention of intercepting packets. Secondly, the packets are consumed by the Blackhole node. Thirdly, the Blackhole nodes can conduct coordinated attacks.

5.2 Greyhole Attack

Grey Hole is a node that can change from behaving correctly to behaving like a black hole so as to avoid detection. Some researchers discussed and proposed a solution to a black hole attack by disabling the ability for intermediate nodes to reply to a Route Reply (RREP) only allowing the destination to reply [15].

5.3 Wormhole Attack

In a wormhole attack, an attacker forwards packets through a out-of-band high quality link and replays those packets at another location in the network [16].

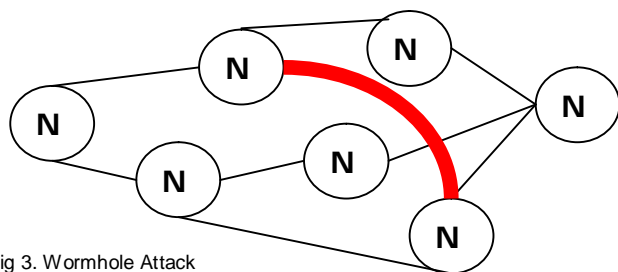


Fig 3. Wormhole Attack

Wormhole attacks depend on a node misrepresenting its location. Hence, location based routing protocols have the potential to prevent wormhole attacks .

5.4 Flooding Attack

In a flooding attack, the attacker targets the target node by flooding in order to congest the network and degrade its performance [17].

5.5 Rushing Attack

In Multicast forwarding a single stream of data is sent to multiple receivers so as to save network bandwidth. A rushing attacker exploits duplicate suppression mechanism by quickly forwarding route discovery packets and hence gaining access to the forwarding group. The goal of this attack is to invade into routing paths. The Multicast routing protocols are targeted which reduce routing overheads. The method is to quickly forward route discovery (control) packets by skipping processing or routing steps [18].

5.6 Jellyfish Attack

In a jellyfish attack the attacker first needs to intrude into the multicast forwarding group. Then data packets forwarding are delayed for some amount of time before forwarding them. This results in significantly high end-to-end delay and thus degrades the performance of real applications. It causes a increase in end –end delay [18].

5.7 Sybil Attack

In a Sybil attack a large numbers of shadow identities are generated and controlled by the malicious parties. Each radio represents a single individual. However the broadcast nature of radio allows a single node to pretend to be many nodes simultaneously by using many different addresses while transmitting [18].

6 Transport Layer Attacks

Transport layer attacks may occur during authentication and securing end-to-end communications which can be rectified through data encryption [9].

The transport layer is responsible for managing end-to-end connections. Two possible attacks in this layer, flooding and desynchronization as in case of sensor networks .

6.1 Flooding

By repeatedly make new connection requests until the resources required by each connection may be exhausted, may reach a maximum limit or may lead to memory exhaustion through flooding. In either case, further legitimate requests will be ignored. One proposed solution to this problem is to require that each connecting client demonstrate its commitment to the connection via the solving of a puzzle. The idea is that a connecting client will not needlessly waste its resources creating unnecessary connections. Since an attacker does not likely have infinite resources, it will be impossible for them to create new connections fast enough to cause resource starvation on the serving node. Although these puzzles do include processing overhead, this technique is still more desirable than excessive communication [20].

6.2 Desynchronization

Disruption of an existing connection is referred to as Desynchronization [19]. An attacker may, for example, repeatedly spoof messages to an end host causing that host to request the retransmission of missed frames. This degrades the ability of the end hosts to successfully send and receive data.

7 Session Layer Attacks

In Session Layer attack an attacker attempts to hijack an established TCP session between two computers by guessing the sequence numbers and taking out one of the user. The counterattacks can be using encryption techniques, limiting incoming traffic etc.

8 Application layer Attacks

Application Layer attacks involve viruses, worms, malicious codes, application abuses etc [9]. When data being transmitted is unencrypted, it is vulnerable to sniffing as well as attacks against applications. Several additional features must be implemented within network equipment to ensure against sniffing of data packets as shown in the figure to provide additional security [21].

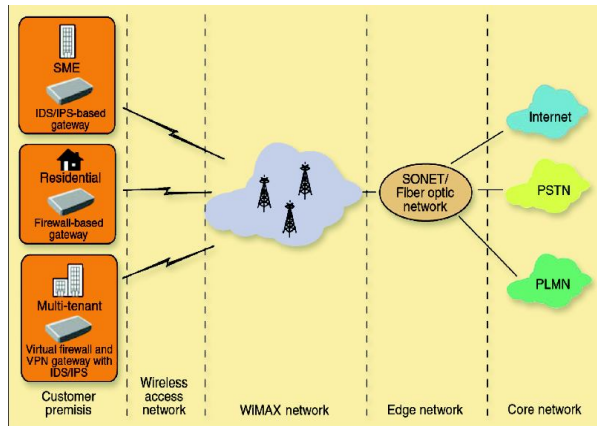


Fig 4. A typical security infrastructure
 Source : [21]

9 Future Work

Future work involves the study of the Identity attacks in Wireless networks. Experimental studies will be conducted in different Wireless Technologies studying the effect of Identity attacks in different scenarios. Possible techniques to identify and counterattack such attacks will be suggested.

References

- [1] S. Vasudevan, K. Papagiannaki, C. Diot, J. Kurose and D. Towsley, "Facilitating Access Point selection in IEEE 802.11 wireless networks", Proceedings of 5th ACM Conference on Internet Measurement, 2005.
- [2] S. Khan, K.Loo, T. Naeem, M. Hhan, " Denial of Service Attacks and Challenges in Broadband Wireless Networks ", IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.7, July 2008.
- [3] V. Gunasekaran, F. Harmantzis, "Emerging wireless technologies for developing countries ", Technology in Society 29 23–42, 2009.
- [4] C. Karlof and D.Wagner, "Secure routing in wireless sensor networks: Attacks and countermeasures", Proceedings of 1st IEEE International Workshop on Sensor Network Protocols and Applications, 2003.
- [5] J. Haartsen, "BLUETOOTH—"The universal radio interface for ad hoc, wireless connectivity" , Ericsson Review No. 3, 1998.
- [6] A. Pirzada, C. McDonald, "Establishing trust in pure ad-hoc networks", in: CRPIT '04: Proceedings of the 27th Conference on Australasian Computer Science, Australian Computer Society Inc., Darlinghurst, Australia, 2004, pp. 47–54.
- [7] H. Kim, Department of Electrical and Computer Engineering, Stevens Institute of Technology, "IEEE 802.16/WiMAX Security", 2006.
- [8] M. Bogdanoski, P. Latkoski, A. Risteski, Borislav Popovsk, "IEEE 802.16 Security Issues: A Survey", 16th Telecommunication forum, TELFOR 2008.
- [9] H Yang, H Luo, F Ye, S Lu, L Zhang , "Security in mobile ad hoc networks: challenges and solutions", IEEE Wireless Communications, 2004 – Citeseer.
- [10] C. Perkins and E Royer, "Ad Hoc On-Demand Distance Vector Routing", 2nd IEEE Wksp. Mobile Comp. Sys. and Apps., 1999.

- [11] D. Johnson and D. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," *Mobile Computing*, T. Imielinski and H. Korth, Ed., Kluwer, 1996.
- [12] Y. Hu, A. Perrig, and D. Johnson, "Ariadne: A Secure On-demand Routing Protocol for Ad Hoc Networks", *ACM MOBICOM*, 2002.
- [13] M. Zapata, and N. Asokan, "Securing Ad Hoc Routing Protocols," *ACM WiSe*, 2002.
- [14] S. Ramaswamy, H. Fu, M. Sreekantaradhya, J. Dixon ,K. Nygard , "Prevention of Cooperative Black Hole Attack in Wireless Ad Hoc Networks" , In Proceedings of the International Conference on Wireless Networks, Las Vegas, June 2003.
- [15] Ebrahim Mohammed Louis Dargin, Oakland University School of Computer Science and Engineering CSE 681 Information Security, "Routing Protocols Security in Ad Hoc Networks", Citeseer.
- [16] L. Hu, D. Evans, Department of Computer Science , University of Virginia Charlottesville, " Using Directional Antennas to Prevent Wormhole Attacks", *Network and Distributed System Security Symposium (NDSS)*, February 2004.
- [17] S. Khan, K.Loo, T. Naeem, M. Khan , "Denial of Service Attacks and Challenges in Broadband Wireless Networks", IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.7, July 2008.
- [18] V. Palanisamy, P. Annadurai, "Impact of Rushing attack on Multicast in Mobile Ad Hoc Network" , (IJCSIS) International Journal of Computer Science and Information Security, Vol. 4, No. 1 & 2, 2009.
- [19] A. D. Wood and J. A. Stankovic, "Denial of service in sensor networks", *IEEE Computer*, vol. 35, no. 10, pp. 54–62, 2002.
- [20] Y. Wang, G. Atterbury, B. Ramamurthy, "A survey of security issues in wireless sensor networks", IEEE Communications Surveys & Tutorials, vol. 8, no. 2, pp. 2-23, 2006.
- [21] R. Mylavaram, "Security considerations for WiMAX- based converged network", <http://rfdesign.com>, August 2005.
- [22] V. Abel, "Survey of Attacks on Mobile Adhoc Wireless Networks", International Journal on Computer Science and Engineering, ISSN : 0975-3397, Vol. 3 No. 2, 2 Feb 2011.
- [23] V. Abel, A. Mnaouer, "On the Study of the WiMAX Security Threats and Current Solution Trends", Journal of the Caribbean Academy of Sciences, 2010.

Anomalous Hydrogen Production during Photolysis of NaHCO_3 Mixed Water

Muhammad Shahid, Noriah Bidin, Yacoob Mat Daud, Muhammad Talha, M.Inayat ullah

Abstract— Production and enhancement of hydrogen on large scale is a goal towards the revolution of green and cheap energy. Utilization of hydrogen energy has many attractive features, including energy renewability, flexibility, and zero green house gas emissions. In this current research the production and the enhancement of hydrogen from the NaHCO_3 mixed water have been investigated under the action of diode pumped solid state laser with second harmonic of wavelength 532nm. The efficiency of the hydrogen and oxygen yields was found to be greater than the normal Faradic efficiency. The parametric dependence of the yields as a function of laser irradiation time, Laser focusing effect and other parameters of the electrolysis fundamentals were carefully studied.

Index Terms— Photo catalysis, Electrolysis of water, Hydrogen, Laser interaction, Electrical signals, Oxygen.

1. INTRODUCTION

We are at the edge of an era of energy crises. The current energy sources are not able to handle the incoming huge population needs. Hydrogen is used at large scale for production of ammonia, for refining the petroleum and also refining the different metals such as uranium, copper, zinc, tungsten and lead etc. The main source of energy on earth is fossil fuels which cause severe pollutions and cannot last for long time use. Nuclear energy is very expensive and having disposal problems. The other sources such as tidal and wind schemes are not sufficient. The solar, thermal and hydal energy sources are feasible but required a lot of capital. An alternative source is water, which is cheap, clean and everlasting source of global energy

Hydrogen gas can be easily obtained by the electrolysis. However, direct decomposition of water is very difficult in normal condition. The pyrolysis reaction occurs at high temperatures above 3700Co . 1) Anomalous hydrogen generation during plasma electrolysis was already reported. 2-5) Access hydrogen generation by laser induced plasma electrolysis was reported recently. 6-9)

Water in the liquid state has the extremely high absorption coefficient at a wavelength of $2.9\ \mu\text{m}$. 10) The effect of generation of an electric signal, when IR-laser radiation having the power density below the plasma formation threshold interacts with a water surface, was discovered by. 11) The electrical signals induced by lasers were already reported. 12,13) A lot of research has been done on photo catalytic hydrogen production. The photo catalytic splitting of water using semiconductors has been widely studied. Many scientists produce hydrogen from water by using different photo catalysts in water and reported hydrogen by the interaction of lasers. 14-18) In addition to this photolysis of water has been studied using UV light. 19) Solar energy has been used to obtained Hydrogen from water by photo cata-

lytic process. 20) But these methods are not economical and the yields of hydrogen is not to an extent.

Our work on lasers has revealed the important parameters which played a critical role in the enhancement of hydrogen from water by laser. Most of the research work basis on photo catalysis has carried out by flash lamps. A very little work is done by lasers. 21) Since laser light has special properties like monochromatic, coherent, intense and polarize, so it was of great interest to use the laser beams as an excitation source in water. The second parameter is that the most of the work has done on light water, distilled water and heavy water; we have used drinking water for production of hydrogen. We have used NaHCO_3 electrolyte. The diode pumped solid state laser having a green light of wave length 532 nm was used as an irradiation source. We investigated the different parameters of the laser by monitoring the rate of evolved gases i.e. hydrogen and oxygen. We inspected the dependence of hydrogen and oxygen yields as a laser exposure time, the effect of laser beam power and the laser focusing effect.

2. Experimental Setup

A schematic diagram of the hydrogen reactor is shown in Figure 2. The reactor contained a glass made hydrogen fuel cell having dimension 10 inch x 8 inch. Fuel cell contained a window for irradiation of laser, an inlet for water and electrolyte, two outlets for hydrogen and oxygen gasses, an inlet for temperature probe and a D.C power supply model ED-345B. Two electrodes steel and Aluminum were adjusted in the fuel cell. A CCD camera and a computer triggered with fuel cell for grabbing, a multimeter and gas flow meter are arranged with the fuel cell. The diode pumped solid state laser with second harmonics DPSS LYDPG-1 model DPG-2000 having green light of wavelength 532 nm was placed near the fuel cell for irradiation during electroly-

sis of water. The DPSS laser spectrum is shown in figure 1. The drinking water 40 ml mixed with 10mL NaHCO₃. In order to start the electrolysis current was applied by D.C source through steel and aluminum electrodes. The laser beam from diode pumped laser was incident on water through window of the fuel cell. The hydrogen and oxygen produced were measured by gas flow meter. The laser beam power was measured by a power meter model Nova Z01500. The temperature of the water was measured by a Temperature probe thermocouple thermometer and mercury thermometer. The current was measured with the help of multimeter. The entire experimental run time was 90 minutes. The data was recorded after every minute of the run.

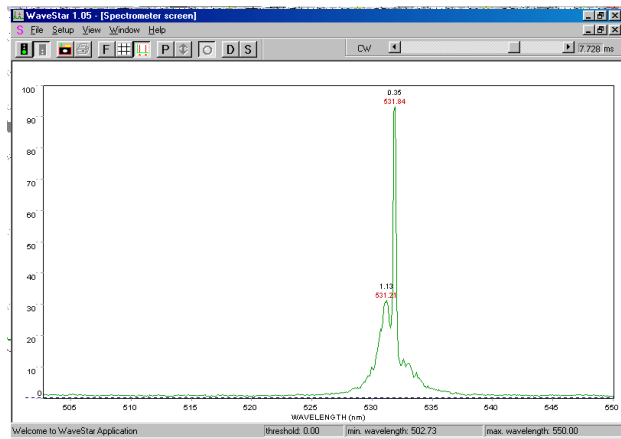


Figure1: DPSS laser spectrum

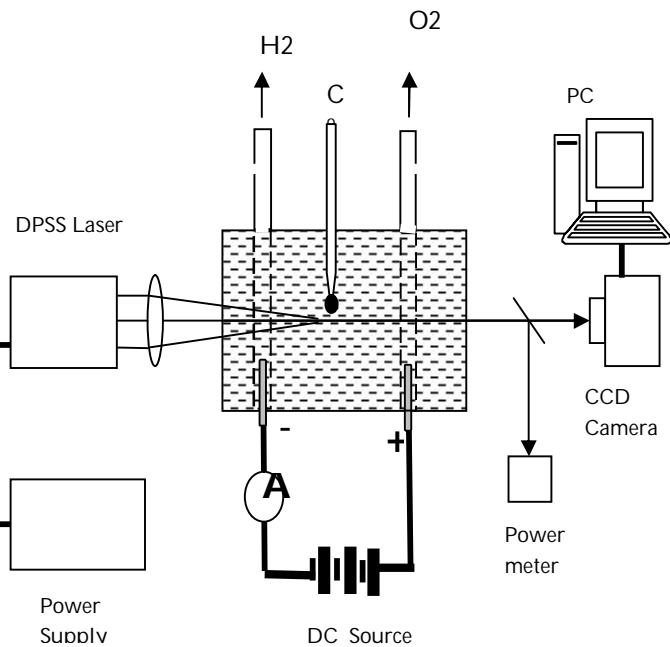


Figure2: Schematic diagram of hydrogen reactor

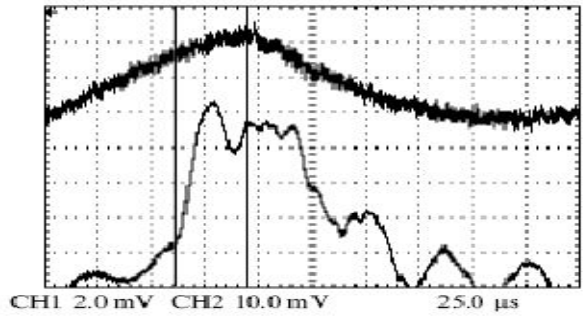


Figure 3. Oscillograms of electrical signal Peaks [7]

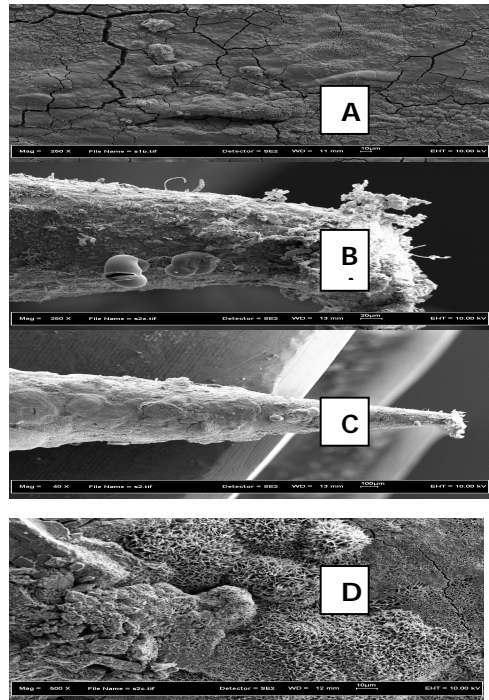
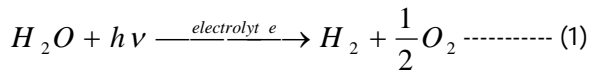


Figure 4. SEM micrographs corrosion on the electrode

Reaction Mechanism



The energy deposited to the water

$$E = VIt + h\nu \text{ ----- (2)}$$

The criteria for splitting water is

$$E \geq E_d$$

and

$$E = E_d + K_H + K_O \text{ -----(3)}$$

Where E is the total energy deposited to the water, E_L is the laser energy, V is the applied D.C voltage, I is the D.C current, t is the current rising time, E_d is the bond dissociation energy of water, KH is the kinetic energy of hydrogen and KO is the kinetic energy of oxygen.

3. Result and discussion

In order to investigate the role of electrolyte as a photo catalyst for water splitting under the influence of laser, various experiments were performed. It has been observed that various important factors affected the yield of hydrogen and oxygen.

- (i) Effect of laser Power
- (ii) Laser focusing effect
- (iii) Effect of Temperature

3.1 Effect of Laser power

The one of the important factors which affected the product yields was Laser power. This effect is shown in figure 5. It has been detected that hydrogen and oxygen yields increased with increase in laser input power. Initially yields were found to be increased linearly with laser power. When the laser power reached at 1 watt a sudden increase was seen in the yields. At this point the hydrogen yield was found to be 0.17cc where as oxygen yield was at 0.1cc. It was due to a sharp electrical signal generated by the laser. This electrical signal peak had enough energy to overcome the bond dissociation energy of water. The hydrogen yield was found to be greater than the oxygen yield. It was all due to the fact a lot of oxygen was used to oxidize the electrodes. So a lot of oxygen was used in corrosion process shown Figure 4.

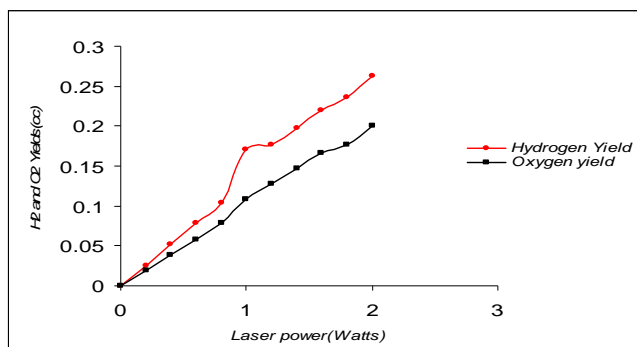


Figure 5. A graph of laser power versus hydrogen and oxygen yields

3.2 Laser focusing effect

It was revealed that the laser focusing effect also affected the yields. The experimental facts showed that when the hydrogen reactor was near the focus of the laser beam, the yield of hydrogen was found to be large. As long as the distance from the focus was increased the production observed to be less. Figure 6 represents this effect. The maximum yield 0.000051 cc was observed at 14 cm distance from the focus where as minimum yield 0.000042 cc was observed at 68 cm from the focal point. It was due to the fact that when reactor was near the focus the intensity of beam was large, at that point powers density became large, so yield of hydrogen found to be large. Similarly when the distance from the focus was increased power density also decreased, so hydrogen yield also observed to be less. The non linear behavior of the curve in the figure 6 shows the moment of the focus point due to stirring of the water. At a distance of 40 cm from the laser focus a sudden change in hydrogen production was observed. This sudden change may be due to external disturbance produced by the stirring effect.

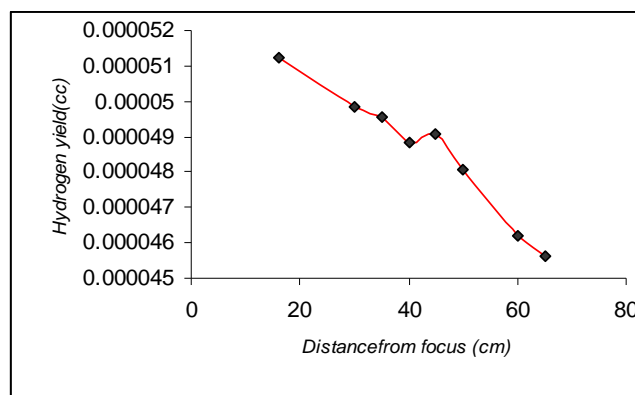


Figure 6. Hydrogen yield versus distance from the laser focus

3.3 Effect of Temperature

The other important factor which affected the product yields was the temperature of the water. It has been observed that temperature of the water rose with time.

It was because of Joule heating effect.

$$H = I^2 R t \text{ (Joules)}$$

Where H is heat dissipation, R is the resistance of water and electrolyte and t is the current rising time. Figure 4 shows this effect. It has also observed that as long as temperature raised yield of hydrogen and oxygen also increased (Figure 7). It was due to the fact that as temperature of water raised, bonding of the water became weak and splitting of water became prominent. The Figure 8 represents the relationship between the laser exposure time

and the efficiency of hydrogen and oxygen yields. It was observed that the efficiency of yields increased rapidly after one minute of the run and reached at 95%. After that efficiency slightly decreased 90% and maintained this value throughout the run of experiment. This efficiency found to be greater than normal faradic efficiency.

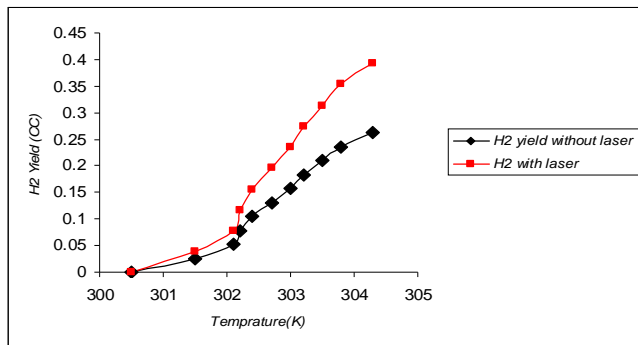


Figure 7. Comparison of hydrogen yields with and without laser

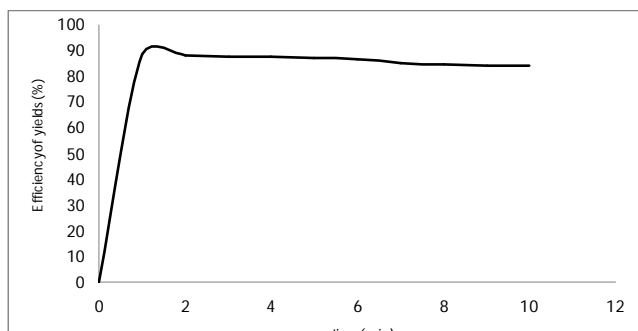


Figure 8. A graph of laser exposure time versus efficiency of yields

3 CONCLUSION

The experimental results revealed that, the diode pumped solid state laser with second harmonics having a green light of wave length 532nm is highly efficient in photo splitting of water into hydrogen and oxygen during plasma electrolysis of NaHCO₃ water. The laser power, focusing effect and temperature of the water have a significant role in enhancement of the hydrogen production photolysis of water.

REFERENCES

- [1] T. Mizuno, T. Ohmori and A. Akimoto "Generation of Heat and Products during Plasma Electrolysis", Proc. ICCF-10, (2003).
- [2] T. Ohmori and T. Mizuno, "Strong Excess Energy Evolution, New Elements Production, and Electromagnetic Wave and/or Neutron Emission in the Light Water Electrolysis with a Tungsten Cathode,"

Proc. ICCF-7, Vancouver, Canada, 279, (2000).

- [3] Tadahiko Mizuno, Tadayoshi Ohmori, Tadashi Akimoto, Akito Takahashi, "Production of heat during Plasma Electrolysis in Liquid", J. Appl. Phys., Vol.39, No.10 (2000).
- [4] N.N. Il'ichev, L.A. Kulevsky, P.P. Pashinin "Photovoltaic effect in water induced by a 2.92- μm Cr³⁺: Yb³⁺: Ho³⁺: YSGG laser" Quantum Electronics 35(10) 959-961 (2005).
- [5] R. Mills, M. Nansteel and P. Ray" Water bath calorimetric study of excess heat generation in "resonant transfer" plasmas " Plasma Physics, 69, 131 (2003).
- [6] Muhammad Shahid, Noriah Bidin, Yacoob Mat" Access Hydrogen production by photolysis of K₂CO₃ mixed water." Proc. FSPGS2010, Malaysia, 179, (2010).
- [7] Muhammad Shahid, Noriah Bidin, Yacoob Mat, "Access hydrogen generation by laser ablation during plasma electrolysis of drinking water." Proc. IGCEH2010, Malaysia, (2010).
- [8] Muhammad Shahid, Noriah Bidin, Yacoob Mat" Novel hydrogen production by laser induced electrical signals during plasma electrolysis of NaOH mixed water." Proc. of PGSRET, Islamabad, Pakistan, Pakistan, (2010).
- [9] Muhammad Shahid, Noriah Bidin, Yacoob Mat, "Laser induced photocatalysis of electrolyte during plasma electrolysis of drinking water." Proc. ICSM2010, Malaysia, (2010).
- [10] Ilyichev N.N. et al. Kvantovaya Elektron., 35 (10), 959 (2005) Quantum Electron, 35 (10), 959 (2005).
- [11] S.A. Andreev, S.Yu. Kazantsev, I.G. Kononov, P.P. Pashinin, K.N. Firsov," Temporal structure of an electric signal produced upon interaction of radiation from a HF laser with the bottom surface of a water column" Quantum Electronics 39 (2) 179 -184 (2009).
- [12] Daniela Bertuccelli and Héctor F. Ranea-Sandoval "Perturbations of Conduction in Liquids by Pulsed Laser-Generated Plasma" IEEE Journal of Quantum electronics, Vol. 37, NO. 7, July 2001.
- [13] Miyuki Ikeda, "Photocatalytic hydrogen production enhanced by laser ablation in water-methanol mixture containing titanium(IV) oxide and graphite silica", catalysis communications 9 (2008) 1329-1333.
- [14] Byeong Sub kwak, "Enhanced hydrogen production from Methanol/water photo-splitting in TiO₂ including Pd component" Bull Korean chem. Soc, Vol.30, No5, (2009).
- [15] M.A Gondal "Production of hydrogen and oxygen by water splitting using laser induced photo-catalysis over Fe₂O₃." Applied catalysis A; general 286(2004) 159-167.
- [16] C. A. Sacchi, "Laser-induced electric breakdown in water," J. Opt. Soc. Amer., vol. B8, pp. 337-345, (1991).
- [17] A. De Giacomo et al, spectroscopic "investigation of laser-water interaction beyond the breakdown threshold energy" CNR. IMIP Sec Bari, via Amendola 122/D, 70126 Bari, Italy (2007).
- [18] F. Andrew Frame, a Elizabeth C. Carroll, a Delmar S. Larsen, a Michael Sarahan, a Nigel D. Browning b and Frank E. Osterloh, "First demonstration of CdSe as a photocatalyst for hydrogen evolution from water under UV and visible light." DOI: 10.1039/b718796c; Berkeley, CA, USA); (2008).
- [19] Ke, stutis Juodkazis^{1,6}, Jurga Juodkazyt¹ "Photoelectrolysis of water: Solar hydrogen-achievements and perspectives" Optical Society of America (2010).
- [20] S.Sino, T.A. Yamamoto, R. Fujimoto "Hydrogen evaluation from water dispersing nano particles irradiated with gamma rays/size effect and dose rate effect." Scripta matter 44(2001) 1709-1712.

- [21] Kia zeng, dongka zhang "Recent progress in alkaline water electrolysis for hydrogen production and applications"progress in energy and combustion science (2009).

Peripheral Interface Controller based the Display Unit of Remote Display System

May Thwe Oo

Abstract— This paper expresses about how to construct the peripheral interface controller (PIC) based the display unit of the remote display system. The remote display system can be used to display the token number that is to know the people. It is also intended for use in clinic, hospitals, bank, and etc. In this research, the peripheral interface controller based remote display system is used for displaying the number and the character. The remote display system consists of two portions: display unit and console unit. The display unit of remote display system contains the display controller, three seven-segment Light Emitting Diode (LEDs), Diode matrix, Category display LEDs and DSUB9 connector. The display controller is controlled by the microcontroller PIC16F873. It controls to display the token numbers. And then it can control at the diode matrix to display the three kinds of character such as A, B, and C. The three numbers of seven-segment LEDs will display the token number from one to 999. The diode matrix helps to display the category display LEDs. The category display LEDs will display one kind of characters. In the research work, the category display LEDs can be displayed only three kinds of characters. For the display unit, the DSUB9 connector applies the data that is from the console unit of the remote display system. In this research work, the display unit works as the receiver the console unit works as the transmitter in remote display system. This paper explains about the design, construction, testing and result of a remote display system.

Index Terms— Diode Matrix, Display Unit, Light Emitting Diode, Peripheral Interface Controller, Remote Display System.

1 INTRODUCTION

THIS display system is very useful in many applications. To advertise, to express the number, the remote display system can be used. Because of the token number can be expressed by using the display system. The token number may be used in banks, financial institutions, clinics and etc. The display unit of the remote display system can be controlled by the microcontroller PIC16F873. The token number is displayed by the large-sized seven-segment Light Emitting Diodes (LEDs). The category display LED can be expressed the three characters. This research consists of two modules; one for the construction of hardware circuit (including interfacing circuit and LED matrix circuit) and other for how to control the process by using peripheral interface controller (PIC). In this research, the hardware controlling program is Assembly programming language. The circuit components consist of resistors, transistors, Integrated Circuit (ICs), LEDs and other accessories. The remote display system consists of two parts. They are console unit and display unit. The display unit is used to display the numbers. That number is needed to know the people. That numbers may be the token number for using many public places. For examples, the token number may be used in the banks and financial institutions, calling the patients one by one at the clinics, calling the candidates at the interviewing and etc. There are many ways to display the number.

2 RELATED WORKS

2.1 PC Basted Token Number Display

An increasing number of banks are computerizing their operations. For actual field use, the display with target LED arrays must be replaced. No other hardware is needed, except the displays and the buffers. The absence of any seven segment decoder chip like 7447 is to be noticed. Because of all the decoding is done entirely by software. Time division multiplexing is another important concept. The four numbers of seven-segment displays share a common data bus. The PC places the first digit on the data bus, and enables only the first seven-segment unit. After a delay of one millisecond, the digit is replaced by the next digit, but this time only the second display unit is enabled. After displaying all the four digits in this way, the first unit is enabling again. The cycle repeats itself over and over. Another feature provided by the software is the ability to suppress leading zeros. Number five is displayed as 5, rather than as 0005. This is usually done by the hardware using the so called 'ripple blanking' facility of the decoder chips. Finally, the computer emits a musical sound whenever there is a change in the displayed number, to attract attention. A regulated 5-volt power supply is recommended [1].

2.2 Digital Bank Token Number Display

Electronic bank token number display systems are quite popular with bank and financial institutions nowadays. These can be used by general practitioners in their clinics to call the patients one by one, and by interviewing committees to call the candidates. In most applications, there

• Miss May Thwe Oo is currently pursuing doctoral degree program in school of computer science and engineering, Beihang University, China, PH-+8613720055716. E-mail:mayooko2009@gmail.com

the area which stores a category display scan position. 'ddisp-p' is to store a seven-segment LED display scan position. 'rcv-p' is the area which counts a received data. And 'r-category' is the workarea to use for the change of the received category data. When making the power ON of the PIC, the instruction is executed from zero address of the program memory. When there is interruption processing, processing is began for the address4. And then it makes each processing jump with the GOTO instruction. The initialization processing is done after turning on. In this processing, all ports of the PORTA are set output mode. And all ports of PORTB are also set to output mode and are used for the segment control of the seven-segment LED. RC0, RC1, RC2, RC3 are set to output mode for the scan of the category display. RC7 pin of PIC16F873 is the only input mode and to receive the transferred data form the console unit.

For the USART, the most of the registers about the setting of a USART are in bank1. Therefore, it is necessary to be careful of the bank designation. The designation of the asynchronous serial communication with 9600bps transmission speed is done. This receiving interrupting occurs when the data is received in the receiving buffer. When the initialization of the USART ends, the receiving operation is immediately done. When the initialization of interruption will start, Global Interrupt Enable bit (GIE), Peripheral Interrupt Enable bit (PEIE), and Timer0 Overflow Interrupt Flag bit (TOIF) are set. To use the interruption of transmission complete, PEIE must to set. The initialization processing is ended. Then, it waits for the interruption only.

As the main processing, it repeats the execution of the same address. In the interruption process subroutine, two kinds of interruptions are used. They are interruption by the time-out of TMR0 and the interruption when the data is received. The interruption of TMR0 is identified by the TOIF bit of the INTCON register and the data receive interruption is identified by the RCIF bit of the PIR1 register.. The RETFIE instruction is executed at end of the interruption processing. It becomes the interruption possible condition.

3.3 Software Flowchart and Processing of LED Control

The interruption occurs every two milliseconds with TMR0. The interruption flag of TOIF should be cleared first. If TOIF bit is not cleared, the interruption occurs without waiting the desired time. Therefore, the setting of the timer value of TMR0 is needed. As for the control of the category display, one row is done every time it interrupts in the two milliseconds. The specification of the row is done by the portC. As for the seven-segment LED control, a three digit figure is displayed by the seven-segment LED. The one digit is controlled every two milliseconds about these LEDs. The digit specification of seven-segments is done by RA0, RA1, and RA2. And RA3, RA4, RA5 of the PIC16F873 is used for the category display

specification. For displaying the seven-segment LEDs, RA0, RA1, RA2 must be rewritten while d\saving this value.

The following flowchart shows the LED control process.

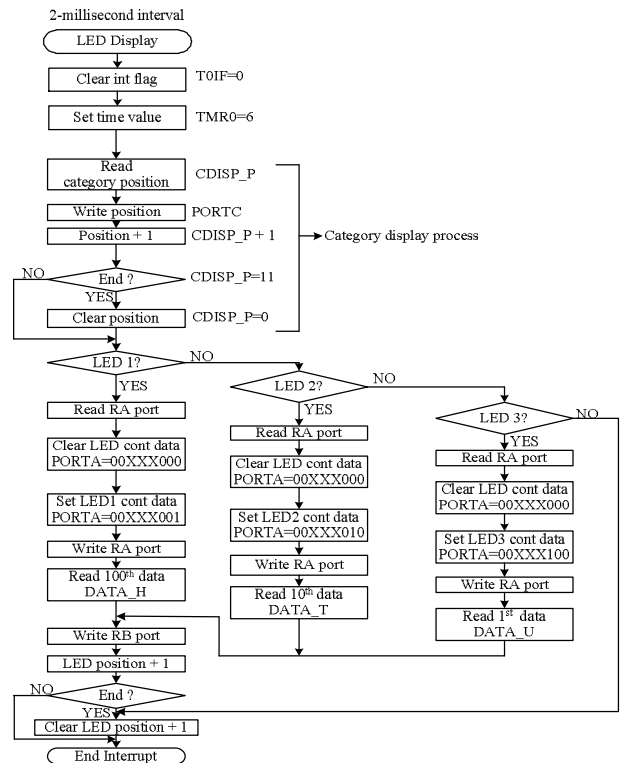


Fig. 4. Software flowchart for subroutine of LED control process.

Therefore, it makes data "0" except the category specification by the AND. By using the AND, the result becomes "1" only when both data is "1". If the bit that fixing value in the instruction is "1", the contents don't change. When the bit that is fixing value is "0" always becomes "0". The digit specification data of the seven-segment LED is set to RA0-2. Now, the calculation of the OR is used. When either of the data is "1", the result becomes "1". If the contents of the bit that fixing value in the instruction is "0", the contents do not change. When the bit that is the fixing value is "1", the contents always becomes "1". When the digit specification is ended, the segment data will be to display. The segment data to display is written in the portB. It reads the content in the received data area of each of the digits and writes in the portB. It makes writing processing to the portB common and a processing step is reduced.

3.4 Software Flowchart for Data Receive Process

The following flowchart is the software flowchart for sub-

routine of data receives process.

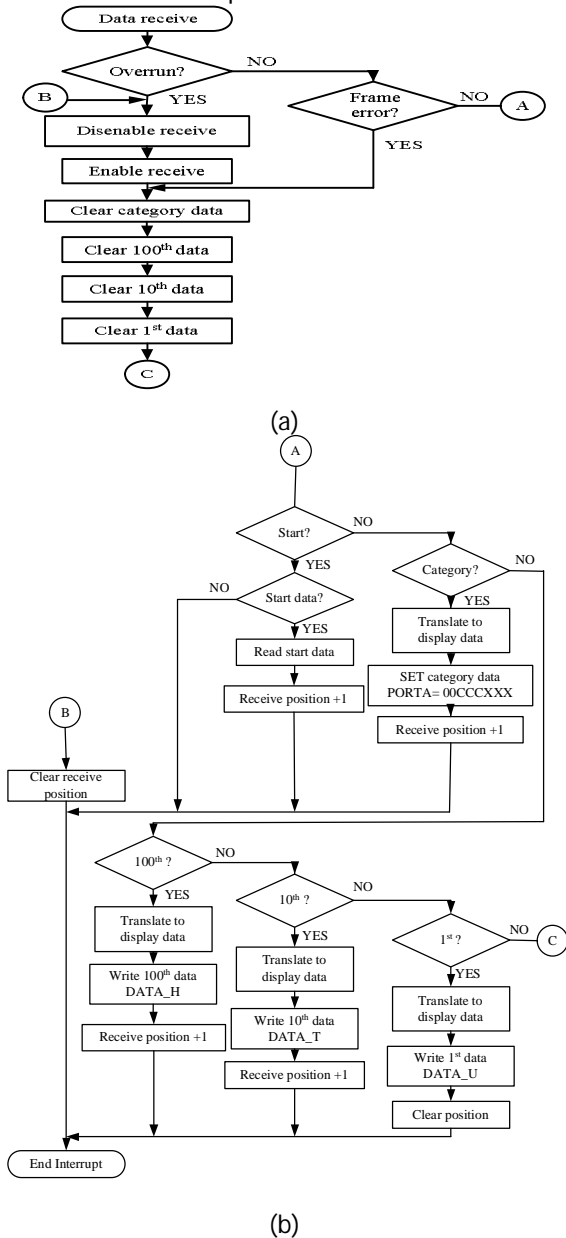


Fig. 5. Software flowchart for subroutine of data receive process: (a) for checking the overrun error and frame error and (b) for checking the category and LED.

The data receive process makes the data receiving from the console unit. This process is started by the data receive interruption by the Receive Interrupt Flag (RCIF) bit of the PIR1 register. It is different from the other interruption display bit and to clear RCIF bit by the software is not necessary. It is cleared by the hardware in reading the data which was received to RCREG. When the data is received, the overrun error must be checked. If an overrun condition occurs, the Continuous Receive Enable (CREN) bit of Receive Status and Control Register (RCSTA) must be cleared. Then the CREN bit is set and

the received data is cleared. When the overrun occurs, it stops the receiver and it must be started once again. This process must be always done when the overrun occurs.

It clears all display data after that to find that the overrun occurred in the process. After the overrun error is checked, the frame error is checked. The frame is from the start bit to the stop bit. When a stop bit is not detected after detecting a start bit, the frame error occurs. If a normal frame is received, the error passes away. When the frame error occurs, it clears display only. If the frame error occurs, the Framing Error (FERR) bit of the Receive Status and Control Register (RCSTA) register is set. And then the category data, 100th data, 10th data and 1st data are cleared. The received position is almost cleared. The interrupt is ended.

After the frame error is checked, the start process is checked. If the frame error does not occur, the start process may be occurred. The category, 100th, 10th and 1st data is sent sequentially from the console unit to the display unit. The information which shows the kind of the data isn't included in each data. The kind of the data is decided in order to receive. The console unit sends start data first and transmits category, 100th, 10th, and 1st data continuously in the order. In the receiving process, it waits for the start data first. If the start data bit is received it is read and the received position is set. Then, the received position is incremented and the interrupt is ended. When the data except the start data comes in the start data receiving position, the data is canceled and the incrementation of the receiving position is not done.

The category data is received behind the start data. After the start data is checked, the category data will be checked. The category data which is sent from the console unit is sent in the form to use in the console unit. The bit configuration and the bit contents of the data that is sent from the console unit to the display unit is different. In the received data, bit four-six is the data which shows a category. Bit six indicates "A", bit five indicates "B", and bit four indicates "C". And, "0" indicates lighting-up and "1" indicates going-out.

If the category data is received, it is checked to be 'A' firstly. As for the display unit, bit three-five indicates a category. Bit three corresponds to "A", bit four corresponds to "B" and bit five corresponds to "C". In the display unit, "0" indicates going-out and "1" indicates lighting-up. The difference is due to the difference of the hardware of each unit. In the number data receive process, the 100th data, the 10th data and the 1st data are received in the order following the category data. The form of the console unit and the form of the display unit are different about these data. The bit position is the same and the meaning is opposite to "0" and "1". In the console unit, '0' is lighting-up and '1' is going-out. In the display unit, "0" is going-out and "1" is lighting-up.

The Exclusion OR is used to reverse "0" and "1". According to Exclusive OR, "0" is output if being a value

with the same data and "1" is output when the value is different. The result may reverses when calculating in Exclusion OR with "1". When the received data is "01001111", the result of execution of Exclusion OR with "11111111" is "10110000". This is the value which reversed "0" and "1" of the received data. A translated value is written in the storage area which corresponds to each of the digits.

When the data receiving stops in the condition that the frame would be received, being normal. The receiving processing has the possibility to make a mistake in the kind of data. When the 100th data is receiving, the communication stopped. And then, when the start data from the console unit is receiving, it receives the start data as 10th data and category data as 1st data. Because it becomes start data waiting condition when 1st data is received, the 100th data, 10th data, and 1st data are canceled. So, the normal receiving is done.

4 HARDWARE IMPLEMENTATION OF DISPLAY UNIT OF REMOTE DISPLAY SYSTEM

The objective of the project is to design the display unit of the remote display system. The display unit is controlled by PIC16F873. The display unit is displayed the number and character that is transferred from the console unit of the remote display system. The display unit of the remote display system consists of three numbers of seven-segment LEDs, diode matrix, RS232C interface, PIC16F873, and four-to-sixteen decoder. The RS232C interface is used to connect between the display unit and the console unit of the remote display system. PIC16F873 controls the three numbers of seven-segment LEDs and diode matrix.

The remote display system consists two portions: Display Unit and Console Unit. This project is to design the display unit of the remote display system. PIC16F873 is used to control both the display unit and the console unit of the remote display system. PORTA of the PIC16F873 is used for output mode to display the number and alphabet. All of PORTB is to display the segments of the seven-segment LEDs. A +5V is used to power of the PIC and +12V is used to the seven-segment LEDs and diode matrix. The diode matrix is used to display three kinds of characters such as A, B, and C.

4.2 Design of Display Unit of the Remote Display System

The display unit of remote display system is controlled by the PIC16F873. There are LED selecting circuit and segment selecting circuit to display the numbers with three numbers of seven-segment LEDs. And there are the row selecting circuit and LED selecting circuit to display the three kinds of characters such as A, B, and C. The RS232C interface is used as the connector between the display unit and the console unit. There is a clock circuit to drive

the PIC16F873. Fig. 6 shows the circuit diagram of the display unit of the remote display system. In the portion of design of display unit of the remote display system, the following circuits are including: the seven-segment LED control circuit, the category display control circuit, the RS232C control circuit, PIC oscillator circuit and power circuit.

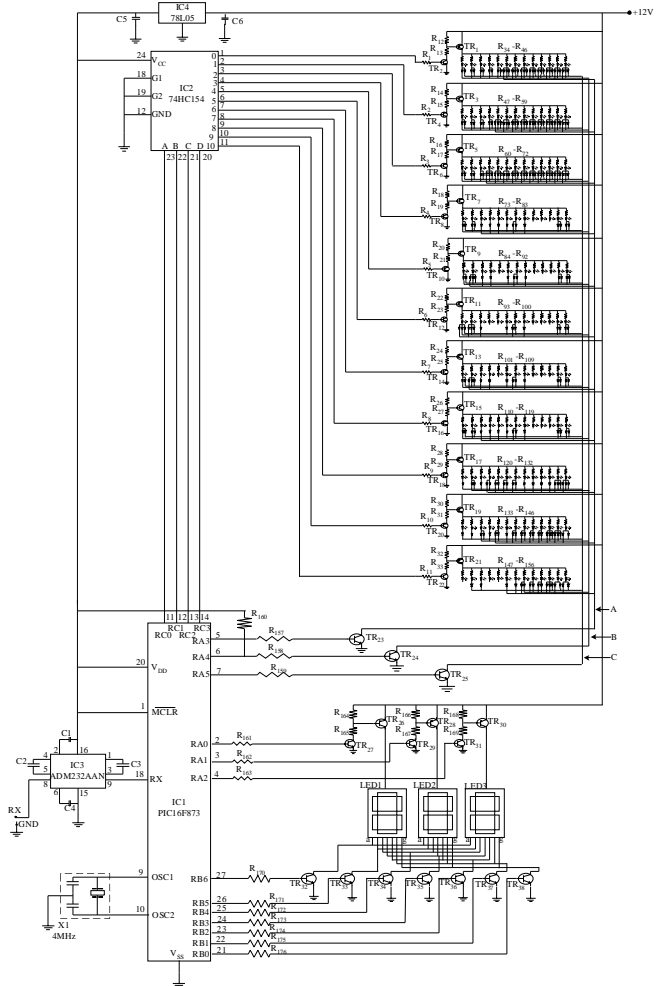


Fig.6. Proposed circuit diagram of the display unit of the remote display system.

4.2.1 Seven-segment LED Control Circuit

In the display unit of the remote display system, there are three number of seven-segment LEDs. It is the I/O port number limitation with PIC and to make an electricity consumption little. Only the circuit which displays 100th data is drawn in the Figure 4.2. The circuit for 10th and 1st is similar too. TR27 and TR26 are the circuit which selects a controlled LED. The drive voltage of the large-sized LED to be using this time is 12V. The drive voltage for one LED is about 2V. As for one segment, five LEDs are connected in series. So, the drive voltage of one segment becomes from 10V to 14V. The LED which was used this time is an anode common type and the anode termin-

al which applies + voltage is common to all the segments. Because an LED selecting circuit is put in the side of the + voltage, PNP type is used for the control transistor. When using NPN-type transistor for this circuit, the emitter of the transistor is connected with the anode terminal of the LED. In the case, the electric current control for the base becomes difficult.

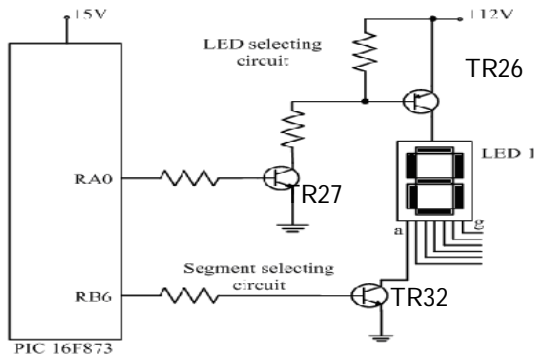


Fig. 7. Schematic diagram of 7-segment LED control circuit.

The TR27 is used for the output of PIC. This is because the voltage which is applied to the I/O port of PIC is limited to +5V. When making TR26 ON, it makes the base voltage of TR26 less than +12V (About 11V). The electric current flows through the base with this. When making TR26 OFF, it makes base voltage +12V. So, about +12V voltage is applied to the base of TR26. The base of TR26 can not be directly driven by PIC. When making RA0 of PIC 0V (0 conditions), TR27 becomes OFF. Therefore, the base electric current of TR26 does not flow and TR26 becomes OFF too. That is, LED1 does not light up. When making RA0 of PIC +5V (1 condition), the base electric current flows through TR27 and TR27 becomes ON. When TR27 becomes ON, the electric current flows through the base of TR26 and TR26 becomes ON condition. By this, +12V are applied on the anode of the LED and the LED becomes the condition about which it is possible to light up.

The segment selecting circuit drives a lit segment. A segment selecting circuit is put between the LED and the grounding. So, the transistor for the control can be directly driven by PIC. In the left figure, a circuit for the "a" segment is drawn. The other segment control circuit is similar too. When making RB6 of PIC 0V (0 conditions), TR32 becomes OFF. Therefore, the electric current does not flow through the "a" segment of the LED and the "a" segment does not light up. The base electric current flows through the base of TR32 when RB6 of PIC is +5V (1 condition). With this, the electric current flows through the "a" segment and the "a" segment lights up. In this research, the common cathode seven-segment LED is used. So, the LED selecting circuit used only one NPN transistor for each seven-segment LED. For the segment selecting circuit, one PNP transistor and one NPN transistor

are used. At the LED selecting circuit, that transistor is directly controlled by the portA of PIC. If the port A of PIC is 0V (0 condition), the transistor will be off and the LED does not light-up. When the portA of PIC is 5V (1 condition), the base electric current flows through the transistor. And the electric current through the LED and the LED will be light-up.

At the segment selecting circuit, the NPN transistor becomes off when the portB of PIC is making 0V (0 condition). So, the base electric current of PNP transistor does not flow and the PNP transistor will off. The electric current does not through the segment of seven-segment LED and the segment does not light-up. If the portB of the PIC is 5V (1 condition), the NPN transistor will be on. The electric current flows through the base of the PNP transistor and that transistor will be on condition. And the electric current flows through the segments and the segment lights up. The PNP transistor is not directly by the PIC.

4.2.2 Category Display Control Circuit

In the display unit, the category is displayed with the diode matrix. Fig. 8 shows the category display control circuit. That consists of the row selecting circuit and the LED selecting circuit. A circuit like seven segments is used about the drive circuit for the category display. A category character is one character but is displayed by the LED matrix. An LED matrix is composed of 11 lines to the direction of the side (Row) and is composed of 13 lines to the longitudinal (Column). Alighting-up control is done every row. The row selecting circuit controls to pass an electric current to the LED of the row which was specified by PIC. Because there are 11 rows, the direct control by PIC is difficult. It is to limit the number of the I/O ports. A control signal is developed from the four-bit signal of PIC to 11 signals by the decoder IC. The operation of the row selecting circuit is similar to the case of seven segments.

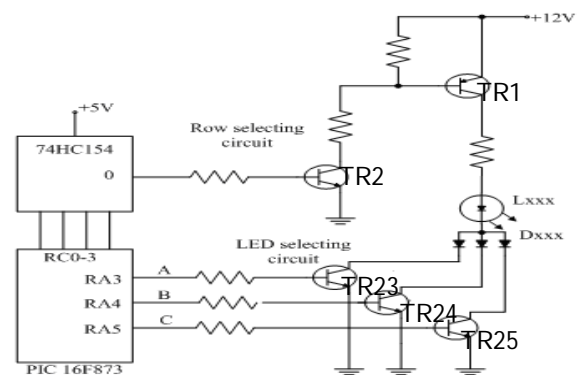


Fig. 8. Schematic diagram of the category display control Circuit.

The LED selecting circuit is the circuit which drives a lit LED in the selected row. The character to display with the circuit this time is three kinds. The kind of the character is controlled by RA3, RA4 and RA5 of PIC. The LED which

is lit up every character is different. When seeing from the LED, there are one which lights up by all character, one which lights up only in case of specific character and so on. It is the diode of the LED selecting circuit that is controlling this. In case of the circuit of the figure on the left, the case of the LED which lights up by all character is shown. When making an LED light up only when TR1 is ON, it makes a diode only the one which is connected with the collector of TR23. It decides a displaying character pattern in the LED matrix and it decides the combination of these diodes every LED.

4.2.3. RS232C Control Circuit

The RS232C interface is used for the information transfer between the console unit and the display unit. This interface is used for an interface with the modem. About $\pm 9V$ is used for signal voltage on the connection cable with the console unit and the long distance transmission is possible comparatively. In the standard, the cable with about 15m length can be used. Actually, it is possible to use for more distance. It depends on the condition of the cable. The +5V signal of PIC16F873 can be directly transmitted without using the RS232C interface if being a short distance. The purpose that this IC is used this time is to attempt to use the RS232C device and to secure a distance. The ADM232AAN has the DC-DC converter which makes the voltage of $\pm 10V$ with +5V power. The external circuit is very simple.

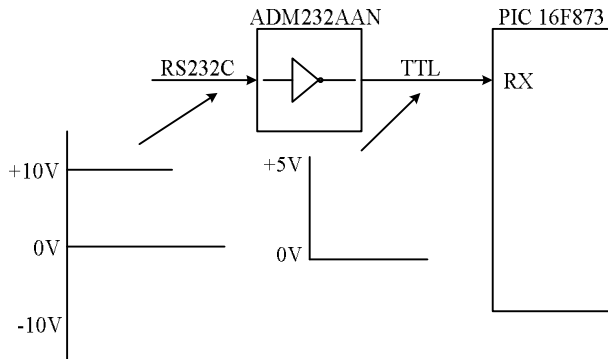


Fig. 9. Schematic diagram of the RS232C control circuit.

4.2.4 PIC Oscillator Circuit

To make do the operation of PIC, a clock generator is needed. An oscillation had within PIC, so, it puts a vibrator outside only. As the vibrator, the crystal oscillator can be used because of the precision oscillation. However, at the circuit this time, a ceramic vibrator (Resonator) is used because it does not need the precision oscillation. The maximum clock frequency of PIC16F873 is upto 20MHz. However, at the equipment this time, because it does not need high-speed operation, the oscillation frequency is 4MHz. A transmission speed is calculated by 4MHz. So, when using this equipment at the clock frequency which is not 4MHz, reconsideration is needed.

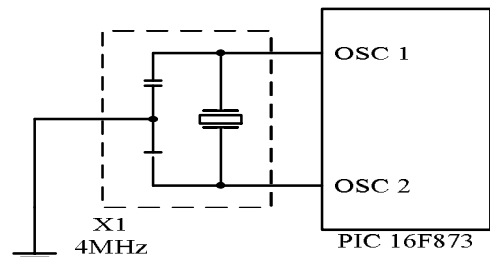


Fig. 10. Schematic Diagram of the PIC Clock Circuit

4.2.5 Power Circuit

The power to operate the display unit is supplied from the console unit through the cable. The DC +12V is made with AC +220V by the switching power circuit. The supplied voltage from the console unit is +12V. The +12V can be used for the LED displaying. The power of PIC needs the +5V. A three-terminal regulator is used for the voltage change from +12V to +5V. The +5V is used for PIC16F873 and decoder IC. The regulator can be used as the 100mA type regulator.

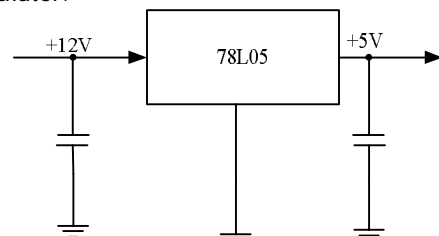


Fig. 11. Schematic Diagram of the Power Circuit

4.3 Test and Result of the Display Unit of the Remote Display System

The peripheral interface controller based the remote display system is shown in Fig. 12. The step-down transformer is used to change from the AC 220V to the 12V. The Fig. 12 is the figure that is before the power supply is off and the data is not being sent from the console unit to the display unit of the remote display system.

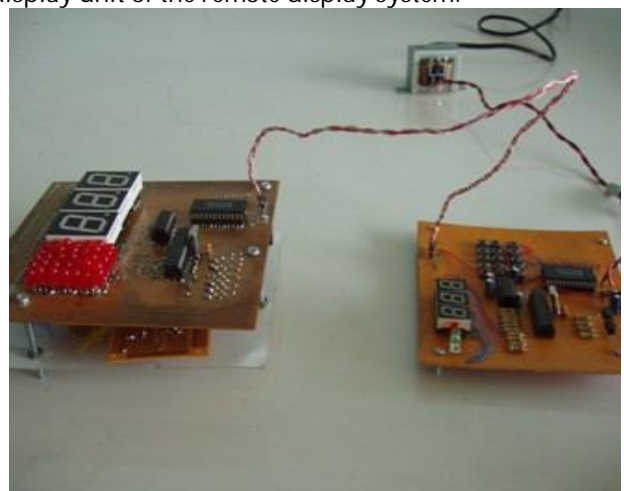


Fig. 12. The PIC based remote display system.

When the number 450 and the character A is sent from the console unit, the data that is sent from the console unit is shown in Fig. 13.



Fig. 13. Testing the receiving data both console unit and display unit in remote display system.

The Fig. 14 shows the display portion of the remote display system.



Fig. 14. The display unit with the data receiving process.

The supply voltage of the display unit is supplied from the console unit of the remote display system. Fig. 15 shows the testing of received voltage level at the display unit.



Fig. 15. Testing the received voltage level at the display unit.

4 CONCLUSION

This research describes about the display unit of the remote display system. The peripheral interface controller based remote display system can be used to inform the people. In this project, the display unit is constructed, one of the nearest, most versatile and most useful LED array, to view alphabetic character and numeric characters. In this research, the three kinds of characters and the three digits numbers are displayed. The PIC16F873 controls the numbers and the characters. The display unit of peripheral interface controller based remote display system is designed to construct using with the devices which can buy easily in the market, easy to use and familiar with a person who interested in electronics field. This system is designed for English language capital letters and numbers. This thesis is not only applied for the educational aid for the beginner to learn the design, programming and development of the applications which use the PIC but also applied for displaying the number in many fields such as clinic, bank and financial institutions, interviewing committees to call the candidates and etc.

As further extension, the remote display system is used 15m length between the console unit and the display unit in this research. For more distance, the various kinds of RS232 interface can be used. In large area of display, it can be used the largest size of seven-segment LED. By this way, the resolution of the display is smooth. Instead of LED, LCD can be used in this research. In this research, the digit display part of the display unit is made to light up by one digit to suppress an electric current. It can be made to light up the digits at the same time. It can be used to hold the display data by latch register (74LS273). The latch register is placed between the peripheral interface controller and the segment selecting circuit. The remote display system can be widely used in many fields that need to give information to users or partners. To develop the display technique and devices, people should

study this technology. The development of this project intended the display unit of the remote display system to be used in various fields.

ACKNOWLEDGMENT

Firstly the author would like to thank her parents for their best wishes to join the Ph.D research. The author would like to express the heart-felt gratitude to Daw Atar Mon for her leadership and advice, Daw Khin Sandar Tun for guidance in her research and U Tun Tun Win for suggestions on how to design the remote token display system. The author greatly expresses her thanks to all persons who will concern to support in preparing this paper and her research.

REFERENCES

- [1] V.Rajarama and Mohan Ingle, "Electronic Project Volume-16", An EFY Enterprise Publication, 1995.
- [2] BEL Application Lab, "Electronic Project Volume-8", An EFY Enterprise Publication, 1987.
- [3] Inoue, S, "Remote Display System", PIC Circuit Gallery, October 2005.
- [4] "Data Sheet PIC16F873 28/40-Pin 8-bit CMOS FLASH Microcontrollers", <<http://www.microchip.com>>
- [5] "+5V-Powered, Multichannel RS-232 Drivers/Receivers." <<http://www.maxim.ic.com/packages>>
- [6] "Data Sheet of 74HC/HC154 4-to-16 Line Decoder/Demultiplexer", September, 2004. <<http://www.philips.com>>
- [7] <http://www.Datasheet4u.com>
- [8] Thomas L. Floyd, "Electronic Devices I, 4th Edition", Prentice Hall International, Inc, 1996.
- [9] Thomas L. Floyd, "Electronic Devices II. 4th Edition", Prentice Hall International, Inc, 1996.
- [10] Thomas E.Kissell, "Industrial Electronics", Printed in the Republic of Singapore, Prentice-Hall International Editions.
- [11] Thomas C.Hayes Paul Horowitz, "Student Manual for the Art of Electronic", Printed in Great Britain at the University Press, Cambridge.
- [12] Nashelsky, L.and Boylestad, R.L., "Electronic Devices and Circuit Theory. 8th Edition", U.S.A. Prentice Hall, Inc. 2002.
- [13] Boylestand, R., "Electronic Devices and Circuit Theory, Fifth Edition", Printed by Prentice-Hall International, Inc.
- [14] <http://www.interq.jp.au.com>
- [15] <http://www.piclist.com>
- [16] <http://www.st.com>
- [17] <http://www.datasheetarchive.com>

Training and Analysis of a Neural Network Model Algorithm

Prof Gouri Patil

Abstract—An algorithm is a set of instruction pattern given in an analytical process of any program/function-ale to achieve desired results. It is a model-programmed action leading to a desired reaction. A neural network is a self-learning mining model algorithm, which aligns/ learns relative to the logic applied in initiation of primary codes of network. Neural network models are the most suitable models in any management system be it business forecast or weather forecast. The paper emphasizes not only on designing, functioning of neural network models but also on the prediction errors in the network associated at every step in the design and function-ale process.

Index Terms— Input , Neural Network, Training, Weights

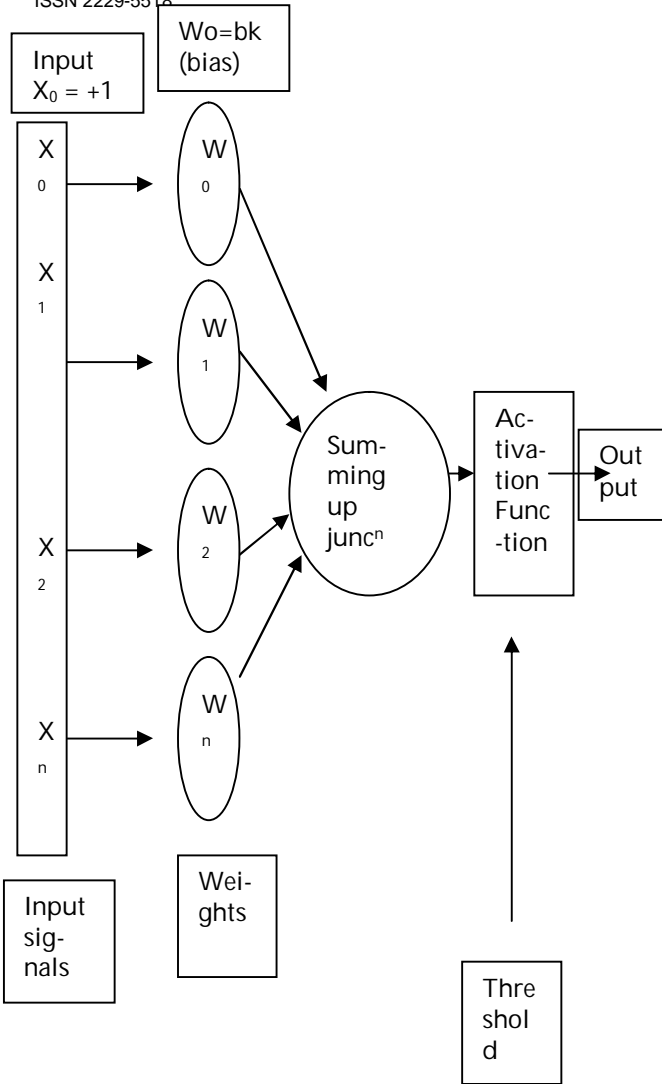
1 NEURAL NETWORK – A MINING MODEL

A Neural Network algorithm could contain multiple networks, depending on the number of columns used for both input and prediction, or that are used only for prediction. The number of networks a single mining model contains depends on the number of parameters connected by the input columns and predictable columns the mining model uses. Neural network function-ale is a mimic of human neural interconnections and memory. Human brain or memory comprises of an average of about ten billion neurons functioning in a network synchronization -- and every single neuron is, on average connected to several other neurons (may be around a thousand) indirectly or directly to the central neural mass, the brain. By way of these connections and interconnections, nerves send and receive messages as packets of energy called as impulses. One very important feature of human neurons is that they don't react immediately to the reception of energy called impulse. Instead, they sum their received energies, and then, they send their own quantities of energy to the other associated neurons only when this sum reaches a certain critical threshold the neurons trigger and respond back as signals or packets of energy called impulse. Brain learns by adjusting the number and strength of these connections and gives a desired response or output in terms of polarization or re-polarization of neurons and difference in the energy levels. Neural networks also work on the same principle as our brains work, they respond to a threshold level of input signal often calculated as weights in the network system.

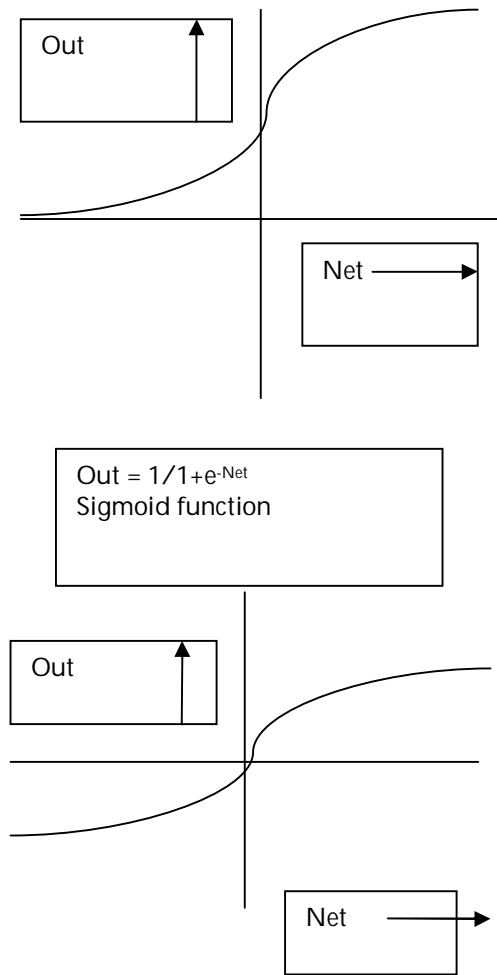
2 INPUT VALUES IN A NEURAL NETWORK MODEL

As all inputs and outputs of neural networks are numeric, primary task in designing a neural network is to define a

transfer function wide enough to accept input in any numeric range and specific enough to give output in the desired range. The range and limit of inputs should be predefined. Saturation/threshold limit for inputs could then be pre-defined, so that inputs in a way are programmed to give the desired outputs or those inputs or functions incompetent of giving the desired outputs are straight away ruled out in a programmed function-ale network model. Networks are built in a manner that they induce a stepwise logical action called the learning rule. Numeric functional values for the inputs are thus predefined and scaled but for some non-numeric nominal values as male or female, yes or no the network functions are divided as different links in the network and graded on an another chain of Numeric functional values as inputs like (0 and 1) and thus branching the network in two simulating paths working on the same predefined threshold value to decide the final output of the process. The input data for training is the fundamental factor of a neural network, as it gives required information to "discover" the optimal threshold operating point and the result or output of the network. If the output of the network is known and weighed inputs are adjusted to reach the desired output then the network training is called supervised neural network. When the output is not defined or known, as in the case of sale or stock prediction the network is an unsupervised network, which is generally programmed to trigger for all hit value of the input. In case of unsupervised networks for hidden neurons, input values of the network may adjust and interpret differentially in different situations and the network ripens or learns based on the adjustments made in the input values of hidden neural layer.



additional benefit of having an extremely simple derivative function for back propagating errors through the neural network.



3 FUNCTION-ALE OF A NEURAL NETWORK MODEL

The adjustments made in the input values of the hidden neural layer of an unsupervised network is nothing but calculations in neural network in terms of adjustments of their weights at the summing up junction and the decisions made at the activation point by defining an activation function value specific for every condition and value. Decisions of the factors include calculations with loops and functions. Initially some time lapse as a function is used to process the data and "winner" input in the time lapse of the model takes all the credit in the first step, where the neuron with the highest value from the calculation fires and takes a value 1, and all other neurons take the value 0.

It is advisable that the sigmoid curve is used as a transfer function because it has the effect of "squashing" the inputs into the range [0,1]. The sigmoid function has the

Typically the weights in a neural network are initially set to small random values; this represents that the network knows nothing to begin with. As the training process proceeds, these weights would converge to values allowing them to perform a useful computation. Thus it can be said that the neural network begins with knowing nothing and moves on to gain admirable real world application.

Activation function is an important function, which decides the maturation and output of a neural network. Activation functions of the hidden values in the neural network could introduce non linearity and desired maturation.

tion of a neural network otherwise the network would have been just a plain mathematical algorithm without logical application. For feedback or feed forward learning of network, the activation function should be differentiable as it helps in most of the learning curves for training a neural network like the bounded sigmoid function, the logistic tanh function with positive and negative values or the Gaussian function. Almost all of these nonlinear activation function conditions assure better numerical conditioning and induced learning.

Networks with threshold limits with out activation functions are difficult to train as they are programmed for step wise constant rise or fall of weight. Where as a sigmoid activation functions with threshold limits makes a small change in the input weight produces change in the output and also makes it possible to predict whether the change in the input weight is good or bad.

In the activation function training, Numerical condition is one of the most fundamental and important concepts of the algorithm, it is very important that the activation function of a network algorithm is a predefined numeric condition. Numerical condition affects the speed and accuracy of most numerical algorithms. Numerical condition is especially important in the study of neural networks because ill-conditioning is a common cause of slow and inaccurate results from many network algorithms.

Numerical condition is mostly decided based on condition number of the input value, which for a neural network is the ratio of the largest and smallest eigenvalues of the Hessian matrix. The eigenvalues of inputs are the squares of the singular values of the primary input and Hessian matrix is the matrix of second-order partial derivatives of the error function with respect to the weights and biases.

4 MODEL NEURAL NETWORK FOR A SAMPLE SIZE OF 20 INDIVIDUALS TO TEST RISK TO DIABETIS.

Type 2 diabetes, is a non-neonatal kind of diabetes that develops in the later stages of life due varied reasons, is far more severe and chronic than type 1 or neonatal diabetes in its destructive metabolic effects on the body. An associated gland called pancreas in our body synthesis hormone called insulin required to metabolize, breakdown and capture glucose for every cell of the body to synthesize energy in the form of energy rich molecule called ATP (Adenosine tri-phosphate). Metabolism of glucose is an indispensable process for the body as this metabolism generates energy for the cell at micro and body at the macro level. There are varied factors response for the onset of this metabolic disorder called diabetes, control of some of these factors could induce synthesis of

insulin in correct amount and at required times so that body learns to metabolize efficiently with inefficient or week pancreas. Ageing is one of the prime factor for all metabolic disorders and so it goes for diabetes, our pancreas ages right along with us, and doesn't pump adequate levels of insulin as efficiently as it did when we were younger. Also, as our cells age, they become more resistant to insulin (the carrier of glucose for the cells) as well. Modern sedentary lifestyle is damaging our healths and is a prime responsive factor for growing obesity problems." being obese or overweight is one of the prime factors for increasing level of glucose in the blood causing diabetes". Obesity increases fat cells in the body and fat cells lack insulin or glucose receptors compared to muscle cells, thus increasing fat in the cells is an indirect call to diabetes, exercising and reducing fat in the cells can act as an alert to stay away from this disorder. Eating less of fat and enough fibre and complex carbohydrates compared to simple carbohydrates could contribute to reduce the risk of diabetes. A survey research on 20 individuals for risk to diabetes was conducted using only 4 non-clinical parameters. A supervised network was built and trained to give an output equivalent to associated risk to diabetes with a predefined threshold limit to reach safe non diabetic level with maximum iterations possible at the activation function to reach the out put, which of course was not fixed but was predefined to reach to zero risk to diabetes.

Sample data collected

No	Age	Stress level/ Kind of work	BMI	Obesity
1	30	Stressful labour	30	Over Weight
2	24	Sedentary work	29	Over weight
3	23	Sedentary work	32	Obese
4	35	Sedentary work	40	Highly obese
5	40	Minimum work	33	Obese
6	45	Minimum work	34	Obese
7	43	Minimum work	22	fit
8	50	Minimum work	26	Slightly over weight
9	55	Maximum work	20	fit
10	58	Maximum work	18	fit
11	37	Maximum work	19	fit
12	38	Minimum work	30	Obese
13	27	Maximum work	25	fit
14	29	Stressful labour	17	fit
15	30	Stressful labour	18	fit
16	34	Sedentary work	20	fit
17	38	Sedentary work	22	fit
18	32	Sedentary work	21	fit
19	42	Maximum work	20	fit
20	43	Maximum work	18	Fit

No	Weight Input Parameter	Weight of input	Weight Range of input	Weight of input obesity	Summing Junction	Predefined Up Weight
1	Age	1	(0-20)	1	1	1
2	Age	2	(20-30)	2	2	2
3	Age	3	(30-40)	3	3	3
4	Age	4	(40-60)	4	4	4
5	Stressful labor	1	Stressful labor	1	8	0
6	Maximum work	1	Maximum work	1	9	1
7	Minimum work	1	Minimum work	1	10	2
8	Sedentary work	1	Sedentary work	1	15	3
9	No work	2	No work	2	12	4
10	BMI	2	(1-25)	3	12	1
11	BMI	3	(25-30)	1	9	2
12	BMI	3	(30-35)	2	11	3
13	BMI	2	(35-40)	1	8	4
14	Obesity test	1	Fit	1	8	1
15	Obesity test	1	Slightly over weight	1	8	2
16	Obesity test	1	Over Weight	1	12	3
17	Obesity test	1	Obese	1	5	4
18	Obesity test	1	Highly obese	1	4	5
15	3	0	1	1	5	
16	3	3	1	1	8	
17	3	3	1	1	8	
18	3	3	1	1	8	
19	4	2	1	1	8	
20	4	1	1	1	7	

Predefined Weights for data
Out put = Threshold limit and risk to diabetes

Condition No	Maximum Threshold limit	Risk to Diabetis
1	5	No risk to Diabetis
2	6	Minimum risk to Diabetis
3	7 or more	Acute risk to diabetes

Network training and summing up Junction

Actiavtion function and Iterations to reach the final output from Summing up Junction

No	Summing up Junction	Activation function	Iterations/training discard	Output-Predict how/when; $\sum n^n \approx 5$
	$\sum n^n$	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1/n^2 / n^3 / n^4 - 1/2/3$ such as $\sum n^n \approx 5$		
1	8	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1$	1	regular exercise required
2	9	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1 - 3$	2	Acute risk to diabetes
3	10	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1 - 5$	4	Diabetic
4	15	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1 - 7$	4	Diabetic

5	12	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1 - 3$	4	Acute risk to diabetes
6	12	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1 - 3$	4	Acute risk to diabetes
7	9	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1$	1	No risk
8	11	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1 - 2$	3	Acute risk to diabetes
9	8	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1 + 1$	1	No risk
10	8	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1 + 1$	1	No risk
11	6	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1 + 2$	1	No risk
12	12	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1 - 4$	4	Diabetic
13	5	$\sum n^n = n^1 + n^2 + n^3 + n^4$	0	No risk
14	4	$\sum n^n = n^1 + n^2 + n^3 + n^4 + 1$	-1 ≈ 0	No risk
15	5	$\sum n^n = n^1 + n^2 + n^3 + n^4$	0	No risk
16	8	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1$	1	No risk
17	8	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1$	1	No risk
18	8	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1$	1	No risk
19	8	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1 + 1$	1	No risk
20	7	$\sum n^n = n^1 + n^2 + n^3 + n^4 - n^1 + 2$	1	No risk

This is a very simple prototype of a network, where every layer of inputs leads to an output. There is no secondary training and learning or firing rule developed in this network, however every network begin with as simple learning single neuron like this and mature to most neural and complex forms. As neural links in the network increases complexity in training and analyzing the network also increases, at the latest with improper links, training and analysis most networks often fails. The next section discusses the common causes for failure of maturing networks and ways to combat these issues with defining open/flexible activation function along with threshold limits to fire an output.

5 SOME COMMON CAUSES AND REMEDIES FOR ILL CONDITIONING IN A NETWORK

1. Low coefficient of variation (standard deviation divided by the mean) of input variables causes

great difference in the numeric functions. Such an issue probably could be minimized by subtracting the mean of each input variable from the variable.

2. If the variances among input variables increase measurably, problem can be cured by dividing each input variable by its standard deviation.
3. High correlations among input variables. This problem could be cured by ortho-normalizing the input variables using Gram-Schmidt, SVD, principal components. The orthogonal components in this case should be standardized before taking as input value. Single orthogonal components which fail standardization should be removed from the network as they cannot be trained or accepted in the network. Training them in the long process may cause network failure.
4. If some hidden training units in the network become unavoidable and in the long run, the network may become saturated with many hidden units. This issue could be handled by using weight decay or similar regularization methods on the input-to-hidden weights, at the risk of less accuracy in learning discontinuities or steep areas of the target function.
5. Low coefficient of variation for activation values of a hidden unit.-this problem can often be ameliorated by using a hidden unit activation function also with an output range of (-1,1), such as tanh, instead of (0,1) as in the logistic function.

tific Digital Literature Library

- [7] Genetic Programming in Data Mining for Drug Discovery - Langdon, Barrett (2004) CiteSeer.IST Scientific Digital Literature Library
- [8] Making Indefinite Kernel Learning Practical - Mierswa (2006) (Correct)
- [9] Evolutionary Learning with Kernels: A Generic Solution for Large.. - Mierswa (2006) CiteSeer.IST Scientific Digital Literature Library
- [10] Combining Decision Trees and Neural Networks for Drug.. - Langdon, Barrett, etal. (2002) CiteSeer.IST Scientific Digital Literature Library
- [11] Diversity in Neural Network Ensembles - Gavin Brown To (2003)
- [12] Linear Equality Constraints and Homomorphous Mappings in PSO - Christopher Monson CiteSeer.IST Scientific Digital Literature Library

6 REFERENCES

- [1] Neural Network design, Martin T Hagan, Howard B Demuth, Mark Beale, 1996 Cengage Learning India Pvt Ltd, ISBN-13 978-81-315-0395-9W.-K. Chen, Linear Networks and Systems.
- [2] III-Conditioning in Neural Networks, Warren S. Sarle, SAS Institute Inc., Cary, NC, USA Sep 5, 1999 Copyright 1999 by Warren S. Sarle, Cary, NC, USA
- [3] The Machine Learning Dictionary for COMP9414
- [4] Analytics Consulting, Prasanna Parthasarathy, Management Journal Bharathidasan Institute of Management, Trichy, 324-327
- [5] Comparison of Artificial Neural Network and Logistic Regression Models for Prediction of Mortality in Head Trauma Based on Initial Clinical Data, Bio-Med Central Feb 2005
- [6] R. Roy.. - Furuhashi And Homann, Soft Computing and Industry Recent Applications, CiteSeer.IST Scientific Digital Literature Library

Wireless Sensor Network: A Review on Data Aggregation

Kiran Maraiya, Kamal Kant, Nitin Gupta

Abstract— Data aggregation is very crucial techniques in wireless sensor network. Because with the help of data aggregation we reduce the energy consumption by eliminating redundancy. When wireless sensor network deployed in remote areas or hostile environment. In the wireless sensor network have the most challenging task is a life time so with help of data aggregation we can enhance the lifetime of the network .In this paper we discuss the data aggregation approaches based on the routing protocols, the algorithm in the wireless sensor network. And also discuss the advantages and disadvantages or various performance measures of the data aggregation in the network.

Index Terms— Wireless sensor network, data aggregation, architecture, Network Lifetime, Routing, Tree, Cluster, Base Station.

1 INTRODUCTION

THE wireless sensor network is ad-hoc network. It consists small light weighted wireless nodes called sensor nodes, deployed in physical or environmental condition. And it measured physical parameters such as sound, pressure, temperature, and humidity. These sensor nodes deployed in large or thousand numbers and collaborate to form an ad hoc network capable of reporting to data collection sink (base station). Wireless sensor network have various applications like habitat monitoring, building monitoring, health monitoring, military survival lance and target tracking. However wireless sensor network is a resource constraint if we talk about energy, computation, memory and limited communication capabilities. All sensor nodes in the wireless sensor network are interact with each other or by intermediate sensor nodes.

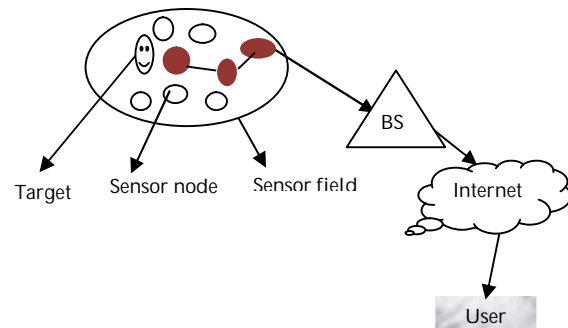


Figure 1 Architecture of the Sensor network

A sensor nodes that generates data, based on its sensing mechanisms observation and transmit sensed data packet to the base station

(sink). This process basically direct transmission since the base station may located very far away from sensor nodes needs.

More energy to transmit data over long distances so that a better technique is to have fewer nodes send data to the base station. These nodes called aggregator nodes and processes called data aggregation in wireless sensor network.

2. CLUSTERING IN WSN

Sensor node are densely deployed in wireless sensor network that means physical environment would produce very similar data in close by sensor node and transmitting such type of data is more or less redundant. So all these facts encourage using some kind of grouping of sensor nodes such that group of sensor node can be combined or compress data together and transmit only compact data. This can reduce localized traffic in individual group and also reduce global data. This grouping process of sensor nodes in a densely deployed large scale sensor node is

- Kiran Maraiya is currently pursuing masters degree program in Computer Science and engineering in NIT Hamirpur, India, PH-+91 9318583266. E-mail: kiran.nitham@gmail.com
- Kamal Kant has joined as Lecturer in ASET, Amity University, Noida (U.P.), India, PH-+919718281158 E-mail: kamalkant25@mail.com
- Nitin Gupta has joined as Assistant Professor in NIT Hamirpur, India, PH-+911972254416 E-mail: nitin@nitham.ac.in

known as clustering. The way of combing data and compress data belonging to a single cluster called data fusion (aggregation).

Issues of clustering in wireless sensor network:-

1. How many sensor nodes should be taken in a single cluster. Selection procedure of cluster head in an individual cluster.

2. Heterogeneity in a network, it means user can put some power full nodes, in term of energy in the network which can behave like cluster head and simple node in a cluster work as a cluster member only.

Many protocols and algorithm have been proposed which deal with each individual issue.

3. DATA AGGREGATION

In typical wireless sensor networks, sensor nodes are usually resource-constrained and battery-limited. In order to save resources and energy, data must be aggregated to avoid overwhelming amounts of traffic in the network. There has been extensive work on data aggregation schemes in sensor networks, The aim of data aggregation is that eliminates redundant data transmission and enhances the lifetime of energy in wireless sensor network. Data aggregation is the process of one or several sensors then collects the detection result from other sensor. The collected data must be processed by sensor to reduce transmission

burden before they are transmitted to the base station or sink. The wireless sensor network has consisted three types of nodes. Simple regular sensor nodes, aggregator node and querier. Regular sensor nodes sense data packet from the environment and send to the aggregator nodes basically these aggregator nodes collect data from multiple sensor nodes of the network, aggregates the data packet using a some aggregation function like sum, average, count, max min and then sends aggregates result to upper aggregator node or the querier node who generate the query.

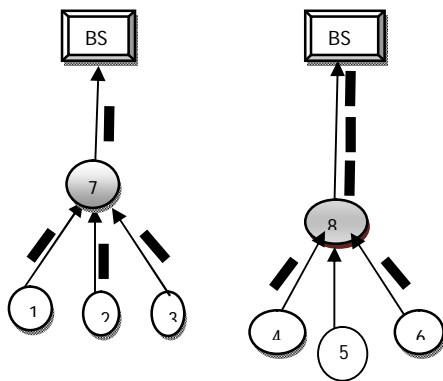


Figure 2 Data aggregation model and Non data aggregation model

It can be the base station or sometimes an external user who has permission to interact with the network. Data transmission between sensor nodes, aggregators and the querier consumes lot of energy in wireless sensor network. Figure 2 contain two models one is data aggregation model and second is non data aggregation model in which sensor nodes 1, 2, 3,4,5,6 are regular nodes that collecting data packet and reporting them back to the upper nodes where sensor nodes 7,8 are aggregators that perform sensing and aggregating at the same time. In this aggregation model 4 data packet travelled within the network and only one data packet is transmitted to the base station (sink). And other non data aggregation model also 4 data packet travelled within the network and all data packets are sent to the base station(sink), means we can say that with the help of data aggregation process we decrease the number of data

packet transmission. And also save energy of the sensor node in the wireless sensor network. With the help of data aggregation we enhance the lifetime of wireless sensor network. Sink have a data packet with energy efficient manner with minimum data latency. So data latency is very important in many applications of wireless sensor network such as environment monitoring, health, monitoring, where the freshness of data is also an important factor. It is critical to develop energy-efficient data-aggregation algorithms so that network lifetime is enhanced. There are several factors which determine the energy efficiency of a sensor network, such as network architecture, the data-aggregation mechanism, and the underlying routing protocol. Wireless sensor network has distributed processing of sensor node data. Data aggregation is the technique. It describes the processing method that is applied on the data received by a sensor node as well as data is to be routed in the entire network. In which reduce energy consumption of the sensor nodes and also reduce the number of transmissions or length of the data packet. Elena Fosolo et al in [5] describe "In network aggregation is the exclusive process of collecting and routing information through a multi hop network. Processing of data packet with the help of intermediate sensor nodes. The objective of this approach is increasing the life time of the network and also reduces resource consumption. There are two type of approach for in network aggregation. With size reduction and without size reduction. In network aggregation with size reduction. It is the process in which combine and compressing the data received by a sensor node from its neighbors in order to reduce the length of data packet to be

sent towards the base station. Example, in some circumstance a node receives two data packets which have a correlated data. In this condition it is useless to send both data packets. Then we apply a function like MAX, AVG, and MIN and again send single data packet to base station. With help of this approach we reduce the number of bit transmitted in the network and also save a lot of energy. In network aggregation without size reduction is defined in the process of data packets received by different neighbors in to a single data packet but without processing the value of data. This process also reduces energy consumption or increase life time of the network.

3.1 Advantage and Disadvantage of Data aggregation in wireless sensor network

Advantage: With the help of data aggregation process we can enhance the robustness and accuracy of information which is obtained by entire network, certain redundancy exists in the data collected from sensor nodes thus data fusion processing is needed to reduce the redundant information. Another advantage is those reduces the traffic load and conserve energy of the sensors. **Disadvantage:** The cluster head means data aggregator nodes send fuse these data to the base station .this cluster head or aggregator node may be attacked by malicious attacker. If a cluster head is compromised, then the base station (sink) cannot be ensure the correctness of the aggregate data that has been send to it. Another drawback is existing systems are several copies of the aggregate result may be sent to the base station (sink) by uncompromised nodes .It increase the power consumed at these nodes.

3.2 Performance measure of data aggregation

There are very important performance measures of data fusion algorithm. These performances are highly dependent on the desired application.

Energy Efficiency: By the data-aggregation scheme, we can increase the functionality of the wireless sensor network. In which every sensor nodes should have spent the same amount of energy in every data gathering round. A data-aggregation scheme is energy efficient if it maximizes the functionality of the network. Network lifetime, data accuracy, and latency are some of the significant performance measures of data-aggregation algorithms. The definitions of these measures are highly dependent on the desired application.

Network lifetime: The network lifetime is defining the number of data fusion rounds. Till the specified percentage of the total nodes dies and the percentage depend on the application .If we talk about some application, simultaneously working of the all the sensor nodes is crucial hence the lifetime of the network is number of round until the first nodes which improves the energy efficiency of nodes and enhance the lifetime of whole network.

Latency: Latency is evaluate data of time delay experiences by system, means data send by sensor nodes and received by base station(sink).basically delay involved in data transmission, routing and data aggregation.

Communication overhead: It evaluates the communication complexity of the network fusion algorithm.

Data accuracy: It is a evaluate of ratio of total number of reading received at the base station (sink) to the total number of generated. There are different types data-aggregation protocols like network architecture based data-aggregation protocols, network-flow-based data-aggregation protocols and quality of service (QoS)-aware data-aggregation protocols designed to guarantee QOS metrics. Here network architecture based protocols are described in detail.

3.3 Impact of data aggregation in wireless sensor network

In this paper we discuss the two main factors that affect the performance of data aggregation methods in wireless sensor network, Such as energy saving and delay. Data aggregation is the process, in which aggregating the data packet coming from the different sources; the number of transmission is reduced. With the help of this process we can save the energy in the network. Delay is the latency connected with aggregation data from closer sources may have to held back at intermediate nodes in order to combine them with data from source that are farther away. Basically aggregation method based on the position of the sources in the network, number of sources and the network topology. If the examine the factors, we consider the two models of the source placement. The event radius (ER) model and random source model [14]. The modelling says us that where the source are clustered near each other or located randomly, significant energy gains are possible with data aggregation. These gains are greatest when the number of sources is large, and when the sources are located relatively close to each other and far from base station. The modelling through, also seems to the suggest that aggregation latency could be non negligible.

4. DATA AGGREGATION APPROACHES IN WIRELESS SENSOR NETWORK

Data aggregation process is performed by specific routing protocol. Our aim is aggregating data to minimize the energy consumption. So sensor nodes should route packets based on the data packet content and choose the next hop in order to promote in network aggregation. Basically routing protocol is divided by the network structure, that's why routing protocols is based on the considered approaches.

4.1 Tree-Based Approach

The tree based approach is defining aggregation from constructing an aggregation tree. The form of tree is minimum spanning tree, sink node consider as a root and source node consider as a leaves. Information flowing of data start from leaves node up to root means sink(base station).Disadvantage of this approach, as we know like wireless sensor network are not free from failure .in case of data packet loss at any level of tree, the data will be lost not only for single level but for whole related sub tree as well. This approach is suitable for designing optimal aggregation techniques'. Madden et al. in [6] data centric protocol know as Tiny aggregation (TAG) approach. The working of TAG is depending on two phases: distributed phase and collection phase.

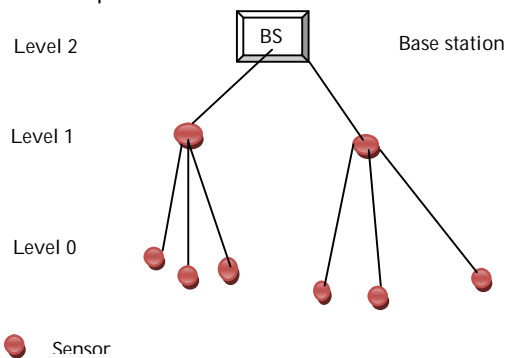


Figure 3 Tree based data aggregation in wireless sensor network

4.2 Cluster-Based Approach

In energy-constrained sensor networks of large size, it is inefficient for sensors to transmit the data directly to the sink. In such scenarios, Cluster based approach is hierarchical approach. In cluster-based approach, whole network is divided in to several clusters. Each cluster has a cluster-head which is selected among cluster members. Cluster-heads do the role of aggregator which aggregate data received from cluster members locally and then transmit the result to base station (sink). Recently, several cluster-based network organization and data-aggregation protocols have been proposed for the wireless sensor network. Figure 4 shows a cluster-based sensor network organization. The cluster heads can communicate with the sink directly via long range transmissions or multi hopping through other cluster heads.

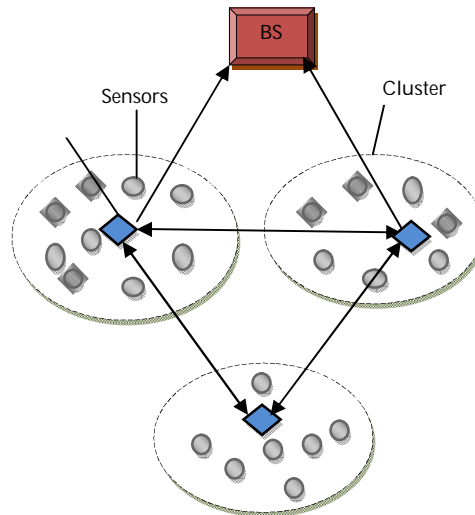


Figure 4 Cluster based sensor network. The arrows indicate wireless communication links

K. Dasgupta et al. in [7] proposed a maximum lifetime data aggregation (MLDA) algorithm which finds data gathering schedule provided location of sensors node and base-station, data packet size, and energy of each sensor node. A data gathering schedule specifies how data packet are collected from sensors node and transmitted to base station for each round. A schedule can be thought of as a collection of aggregation trees. In [4], they proposed heuristic-greedy clustering-based MLDA based on MLDA algorithm. In this they partitioned the network in to cluster and referred each cluster as super-sensor. They then compute maximum lifetime schedule for the super-sensors and then use this schedule to construct aggregation trees for the sensors. W. Choi et al. in [1] present a two-phase clustering (TPC) scheme. Phase I of this scheme creates clusters with a cluster-head and each node within that cluster form a direct connects with cluster-head. Phase I the cluster-head rotation is localized and is done based on the remaining energy level of the sensor nodes which minimize time variance of sensors and this lead to energy saving from unnecessary cluster-head rotation. In phase II, each node within the cluster searches for a neighbor closer than cluster-head which is called data relay point and setup up a data relay link. Now the sensor nodes within a cluster either use direct link or data relay link to send their data to cluster head which is an energy efficient scheme. The data relay point aggregates data at forwarding time to another data relay point or cluster-head. In case of high network density, TPC phase II will setup unnecessary data relay link between neighbors as closely deployed sensor will sense same data and this lead to a waste of energy.

4.3 Multi-path Approach

The drawback of tree based approach is the limited robustness of the system. To overcome this drawback, a new approach was proposed by many researchers .in which sending partially aggregated data to single parent node in aggregation tree, a node could send data over multiple paths. In which each and every node can send data packets to its possibly multiple neighbours. Hence data packet flow from source node to the sink node along multiple path, lot of intermediate node between source node to sink node so aggregation done in every intermediate node. Using this approach we will make the system robust but some extra overhead. The example of this approach like ring topology, where network is divided in to concentric circle with defining level levels according to hop distance from sink.[3]propose a new strategy have both issues : energy efficiency and robustness. In which single path to connect each node to the base station it is energy saving but high risk of link failure. But on the other head multipath approach would require more nodes to participate with consequent waste of energy. Authors present a clever use of multi-path only when there is loss of packet which is implemented by smart caching of data at sensor nodes. Authors also argue that in many practical situation data may be gathered only from a particular region, so they use a different approach that relies on a spanning tree and provides alternative paths only when a malfunctioning is detected. Algorithm adopts a tree-based approach for forwarding packets through the network. In the ideal situation when no failures occur, this is certainly the best choice, as the minimum number of nodes is engaged in the transmission phase. In the presence of link or node failures, the algorithm will discover alternative paths, so as ensure the delivery of as many packets as possible within the time constraints. The problem with this approach is that it may cause the arising of hot spots and nodes along preferred paths will consume their energy resources quickly, possibly causing disconnection in the network.

4.4 Hybrid Approach

Hybrid approach followed between tree, cluster based and multipath scheme. In which the data aggregation structure can adjusted according to specific network situation and to some performance statistics.

Table 1
Routing protocol for Tree, cluster, Multipath and Hybrid approach.

Protocols/algorithms	Tree	Cluster	Multipath	Hybrid
TAG	✓	-	-	-

Directed Diffusion	✓	-	-	-
PEGASIS	✓	-	-	-
DB-MAC	✓	-	-	-
EADAT	✓	-	-	-
LEACH	-	✓	-	-
Cougar	-	✓	-	-
Synopsis	-	-	✓	-
Diffusion	-	-	✓	-
Tributaries and Deltas	-	-	-	✓

5. DATA AGGREGATION FUNCTION IN WIRELESS SENSOR NETWORK.

Many effective type of data aggregation function is needed in wireless sensor network. These functions are closely related to sensor network application. Such as mean quantile, medium, count, average, max, and min.

5.1 Duplicate sensitive and duplicate insensitive

Aggregation function may be average and minimum. If we use average function , it take as a duplicate sensitive and minimum function is take as duplicate insensitive function in wireless sensor network. Data aggregated in the network on that time same data consider multiple times. We can used duplicate function then the final result depends on the number of the times and same value has been considered otherwise aggregation function is said to be duplicate insensitive.

5.2 Lossy and lossless

Data packet can be aggregated with the help of lossy aggregation or by lossless aggregation. Lossy aggregation approach does not follow a perfect reconstruction but lossless aggregation ensures a complete recovery of all individual sensor data at base station (sink) [2].

6. DATA REPRESENTATION IN WIRELESS SENSOR NETWORK.

Data representation is the effective way to representation the data. Wireless sensor network is consisting a large number of small sensor nodes. These are resource constraint, due to limited resource constraint it needs to decide whether to store, compress, discard and transmit

data. All this requirement wants a suitable way to represent the information any type of structure are common to all sensor node in the network.[14]

7. SECURITY ISSUES IN DATA AGGREGATION FOR WIRELESS SENSOR NETWORK

There are two type of securities are require for data aggregation in wireless sensor network, confidentiality and integrity. The basic security issue is data confidentiality, it is protecting the sensitive data transmission and passive attacks, like eavesdropping. If we talk about hostile environment so data confidentiality is mainly used because wireless channel is vulnerable to eavesdropping by cryptography method. The complicated encryption and decryption operations such as modular multiplication. The

8. CONCLUSION

In this paper we present wireless sensor network is consist a large number of sensor node. And these nodes are resource constraint. That's why lifetime of the network is limited so the various approaches or protocol has been proposed for increasing the lifetime of the wireless sensor network. In this paper we discuss the data aggregation are one of the important techniques for enhancing the life time of the network. And

security issues is data integrity with the help of integrity we reduce the compromised sensor source nodes or aggregator nodes from significantly altering the final aggregation value. Sensor node in a sensor network is easily to compromised. Compromised nodes have a capability to modify or discard messages. Method of secure data aggregation: There are two type of method for securing data hop by hop encryption and end to end encryption, both methods follows some step.

1. Encryption process has to be done by sensing nodes in wireless sensor network.
2. Decryption process has to be done by aggregator nodes.
3. After that aggregator nodes aggregates the result and then encrypt the result again.
4. The sink node gets final aggregated result and decrypt it again.

also discuss the various approaches for data aggregation or also discuss the advantage and disadvantages and various performance measures of the data aggregation.

ACKNOWLEDGMENT

This work was supported in part by a grant from NIT Hamirpur (himachal Pradesh).

REFERENCES

- [1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A Survey On Sensor Networks", *IEEE Communications Magazine*, Volume 40, Number 8, pp.102-114, 2002.
- [2] T. Arampatzis, J. Lygeros, and S. Manesis, "A Survey Of Applications Of Wireless Sensors And Wireless Sensor Networks", *In Mediterranean Conference On Control And Automation MED05*, Nicosia, Cyprus, 2005.
- [3] L. Gatani, G. Lo Re, and M. Ortolani, "Robust and Efficient Data Gathering for Wireless Sensor Networks", in *Proceeding of the 39th Hawaii International Conference on System Sciences – 2006*
- [4] K. Dasgupta, K. Kalpakis, and P. Namjoshi, "An Efficient Clustering-based Heuristic for Data Gathering and Aggregation in Sensor Networks", *IEEE 2003*
- [5] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "In-Network Aggregation Techniques for Wireless Sensor Networks: A Survey", *IEEE Wireless communication 2007*.
- [6] S. Madden *et al.*, "TAG: a Tiny Aggregation Service for Ad-hoc Sensor Networks," *OSDI 2002*, Boston, MA, Dec. 2002.
- [7] K. Dasgupta *et al.*, "Maximum Lifetime Data Gathering and Aggregation in Wireless Sensor Networks", *In Proc. of IEEE Networks'02 Conference*, 2002.
- [8] M. Ding, X. Cheng and G. Xue, "Aggregation Tree Construction in Sensor Networks," *2003 IEEE 58th Vehic. Tech. Conf.*, vol. 4, no. 4, Oct. 2003, pp. 2168–72.
- [9] K. Vaidyanathan *et al.*, "Data Aggregation Techniques in Sensor Networks," Technical Report, OSU-CISRC-11/04-TR60, Ohio State University, 2004
- [10] V. Raghunathan, C. Schugers, S Park, and M.B. Srivastava, "Energy-Aware Wireless Microsensor Networks", *IEEE Signal Processing Magazine*, Volume 19, Number 2, pp. 40-50, 2002.
- [11] Lin F.Y.S., Yen H.H., Lin S.P., Wen Y.F.: 'MAC aware energyefficient data-centric routing in wireless sensor networks'. *Proc. IEEE Int. Conf. Commun. (ICC)*, 2006, vol. 8, pp. 3491–3496
- [12] Krishnamachari B., Estrin D., Wicker S.: 'Modeling data-centric routing in wireless sensor networks', USC Computer Engineering Technical Report, CENG 02-14, 2002.
- [13] H.H. Yen, C.L. Lin: 'Integrated channel assignment and data aggregation routing problem in wireless sensor networks', *IET Communications*, 2009, Vol. 3, Iss. 5, pp. 784–793.
- [14] Ramesh Rajagopalan and Pramod K. Varshney, Syracuse University "Data aggregation techniques in sensor network : A survey" "IEEE communication survey & tutorial.4th quarter 2006

Slant Transformation as a Tool for Pre-processing in Image Processing

Nagaraj B Patil, V M Viswanatha , Dr. Sanjay Pande MB

Abstract— The Slantlet Transform (SLT) is a recently developed multiresolution technique especially well-suited for piecewise linear data. The Slantlet transform is an orthogonal Discrete Wavelet Transform (DWT) with 2 zero moments and with improved time localization. It also retains the basic characteristics of the usual filterbank such as octave band characteristic and a scale dilation factor of two. However, the Slantlet transform is based on the principle of designing different filters for different scales unlike iterated filterbank approaches for the DWT. In the proposed system, Slantlet transform is implemented and used in Compression and Denoising of various input images. The performance of Slantlet Transform in terms of Compression Ratio (CR), Reconstruction Ratio (RR) and Peak-Signal-to-Noise-Ratio (PSNR) present in the reconstructed images is evaluated. Simulation results are discussed to demonstrate the effectiveness of the proposed method.

Index Terms—Discrete Wavelet transform, Compression ratio, Data Compression, Peak-signal-to-ratio(PSNR), Coding Inter Pixel, Slantlet Coefficients, Choppy Images.

1 INTRODUCTION

For many decades, scientists wanted more appropriate functions than the sines and cosines, which comprise the basis of Fourier analysis, to approximate choppy images. By their definition, these functions are non-local (and stretch out to infinity). Therefore they do a very poor job in approximating sharp spikes. Wavelets are functions that satisfy certain mathematical requirements and are used in representing data or other functions. This makes wavelets interesting and useful. But with wavelet analysis, approximating functions can be used that are contained neatly in finite domains. Wavelets are well suited for approximating data with sharp discontinuities [1].

The Discrete Wavelet transform (DWT) is usually carried out by filterbank iteration, but for a fixed number of zero moments it does not yield a discrete time basis that is optimal with respect to time localization. The Slantlet transform is an orthogonal DWT with 2 zero moments and with improved time localization. The Slantlet transform has been developed by employing the lengths of the discrete time basis function and their moments as the vehicle in such a way that both time-localization and smoothness properties are achieved. Using Slantlet transform it is possible to design filters of

shorter length while satisfying orthogonality and zero moments condition. The basis function retains the octave-band characteristic. Thus Slantlet transform has been used as a tool in devising an efficient method for compression and denoising of various images.

2. Literature survey

G.K. Kharate, A.A. Ghatol, et.al [2], proposed an algorithm based on decomposition of images using Daubechies wavelet basis. Compression of data is based on reduction of redundancy and irrelevancy. They have proposed an adaptive threshold for quantization, which is based on the type of wavelet, level of decomposition and nature of data.

Sos S. Agaian, Khaled Tourshan, et.al [3] proposes a novel approach for the parameterization of the slantlet transform with the classical slantlet and the Haar transforms are special cases of it is presented. The slantlet transform matrices are constructed first and then the filterbank is derived from them. The parametric slantlet transform performance in image and signal denoising is discussed.

Panda. G, Dash. P K, et.al [4] proposes a novel approach for power quality data compression using the

slantlet is presented and its performance in terms of compression ratio (CR), percentage of energy retained and mean square error present in the reconstructed image is assessed.

Lakhwinder Kaur, Savita Gupta, et.al [5] proposed an adaptive threshold estimation method for image denoising in the wavelet domain based on the generalized Guassian distribution (GGD) modeling of subband coefficients.

Lei Zhang; Bao, P; Xiaolin Wu [8] proposed a wavelet-based multiscale linear minimum mean square-error estimation (LMMSE) scheme for image denoising is proposed, and the determination of the optimal wavelet basis with respect to the proposed scheme is discussed. The overcomplete wavelet expansion (OWE), which is more effective than the orthogonal wavelet transform (OWT) in noise reduction, is used.

3.The Slantlet Transform

The Slantlet filterbank is an orthogonal filter bank for the discrete wavelet transform, where the filters are of shorter support than those of the iterated D_2 filterbank tree. This filterbank retains the desirable characteristics of the usual DWT filterbank.

The Slantlet filter bank shown in Fig 3.1. is generalized as follows. The l -scale filter bank has $2l$ channels. The low-pass filter is to be called $h_l(n)$. The filter adjacent to the lowpass channel is to be called $f_l(n)$. Both $h_l(n)$ & $f_l(n)$ are to be followed by down sampling by 2^l . The remaining $2l - 2$ channels are filtered by $g_i(n)$ & its shifted time-reverse for $i = 1, \dots, l-1$. Each is to be followed by down sampling by 2^{i+1} . $x(n)$ Slantlet filter Coefficients

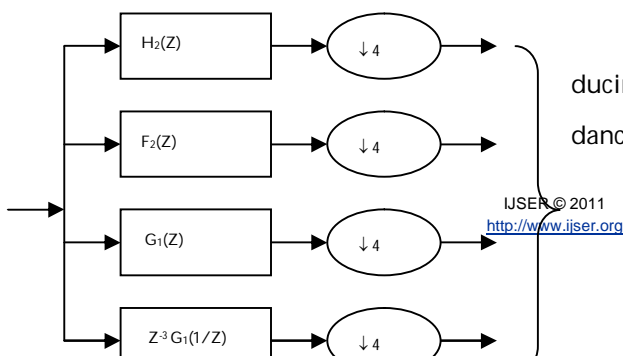


Fig 3.1. Slantlet filterbank

In the Slantlet filterbank, each filter $g_i(n)$ does not appear together with its time reverse. While $h_i(n)$ does not appear with its time reverse, it always appears paired with the filter $f_i(n)$. In addition, note that the l -scale and $(l + 1)$ scale filterbanks have in common the filters $g_i(n)$ for $i = 1, \dots, l-1$ and their time-reversed versions.

The Slantlet filterbank analyzes scale i with the filter $g_i(n)$ of length 2^{i+1} . The characteristics of the Slantlet filterbank:

1. Each filterbank is orthogonal. The filters in the synthesis filter bank are obtained by time reversal of the analysis filter.
2. The scale-dilation factor is 2 for each filterbank.
3. Each filterbank provides a multiresolution decomposition.
4. The time localization is improved with a degradation of frequency selectivity.
5. The Slantlet filters are piecewise linear.

4.Compression

The term data compression refers to the process of reducing the amount of data required to represent a given quantity of information. It is useful both for transmission and storage of information[7].

The encoder of the Fig 4.1 is responsible for reducing the coding interpixel and/or psychovisual redundancy of the input data. In the first stage of the encoding

process the mapper transforms the input data into a format designed to reduce the interpixel redundancies.

The decoder contains only two components: a symbol decoder and an inverse mapper. The percentage of reconstruction ratio can be defined as

Vector norm of the retained SLT Coefficients after threshold

X100

Vector norm of the original SLT coefficients

Compression of data without degradation of data quality is possible because the data contains a high degree of redundancy. The higher the redundancy the higher the achievable compression. The data compression methods can be broadly classified into

1. Lossless compression method
2. Lossy compression method.

Fig4.1.General Block diagram for compression

The second stage, or quantizer block, reduces the accuracy of the mappers output in accordance with a predefined fidelity criterion- attempting to eliminate only psychovisually redundant data. This operation is irreversible and must be omitted when error-free compression is desired. In the third and final stage of process, a symbol coder creates a code (that reduces coding redundancy) for the quantizer output and maps the output in accordance with the code.

If n_1 and n_2 denotes the number of informations carrying units in the original and encoded data respectively, the compression that is achieved can be quantified numerically via the compression ratio

$$C_R = n_1/n_2$$

To view and/or use a compressed (i.e., encoded) data, it must be fed into a decoder, where a reconstructed output data is generated. Error can be defined between f and f^1

$$e = f^1 - f \quad \dots\dots\dots 4.1$$

5.DENOISING

The de-noising objective is to suppress the noise part of the signal s and to recover f while retaining as much as possible the important signal features.

The noisy signal is basically of the form $s(n) = f(n) + \sigma e(n)$, where $e(n)$ is the gaussian white noise $N(0,1)$ and time n is equally spaced.

In the recent years there has been a fair amount of research on wavelet threshold and threshold selection for image de-noising, because wavelet provides an appropriate basis for separating noisy signal from the data signal. Threshold is a simple non-linear technique, which operates on one wavelet coefficient at a time. Replacing the small noisy coefficients by zero and inverse wavelet transform on the result may lead to reconstruction with the essential signal characteristics and with less noise. The Peak-Signal-To-Noise-Ratio(PSNR) for any given data can

be given as, $PSNR=20 \cdot \log_{10}(1256 / (\text{original data-de-noised data}))$.

6.Design Procedure

6.1.Algorithm for signal compression

The design procedure for signal compression contains three steps:

1. Decomposition

The Slantlet transform is used for decomposition and applied to each row and then again applied to the resulting information in each column. The steps used for decomposition

- 1.Find out whether the input signal is power of 2.
- 1.2. Separate the odd and even moment vectors along the length of the input signal.
- 1.3. Since the filters are piecewise linear, each filter can be represented as the sum of a DC and a linear term.
- 1.4. The DC and linear moments at scale i can be computed from the DC and linear moments at the next finer scale (i-1).

The image obtained after 2-D Slantlet transform can be shown as

LL ₃ LH ₃	LH ₂	LH ₁	
HL ₃ HH ₃			
HL ₂	HH ₂		
HL ₁	HH ₁		

Figure 5.3. Decomposed image after 2-D Slantlet Transform

2. Threshold the coefficients

If a pixel in the image has intensity less than the threshold value, the corresponding pixel in the resultant image is set to white. Otherwise, if the pixel is greater than or equal to the threshold intensity, the resulting pixel is set to black. Soft threshold is used for image compression.

3.Reconstruction

Compute wavelet reconstruction based on the modified coefficients. This step uses the inverse Slantlet transform to perform the wavelet reconstruction.

- 3.1. Find out whether the signal is of finite duration (power of 2).
- 3.2. Initialize the moment vectors.

$$\mu_0 = 1:N/2$$

$$\mu_1 = \mu_0$$

- 3.3. Using DC and linear Slantlet coefficients, compute $\mu_0(n;l)$ and $\mu_1(n;l)$

$$m = 2^l;$$

$$\mu_0(n;l) = s(1) / \sqrt{(m)} * s(2) * \sqrt{(3 * (m-1) / (m * (m+1)))};$$

$$\mu_1(n;l) = s(2) * (-2 * \sqrt{(3 / (m * (m^2 - 1))})$$

- 3.4. Then compute $\mu_0(n;i)$ and $\mu_1(n;i)$ for decreasing values of i by updating μ_0 and μ_1 using Slantlet coefficients.

6.2. Algorithm for signal De-noising

Input image is added with noise. For this noisy image, apply Slantlet transform. Steps 1 and 3 of the section 6.1 remains unchanged. But in the step 2, soft threshold is applied for de-noising. Soft threshold is an extension of hard threshold, first setting to zero the elements whose absolute values are lower than the threshold, and then shrinking the nonzero coefficients towards 0. The output of the step3 in section 6.1 is the de-noised version of the original input image.

7. Results and Analysis

Computer simulation study on Compression and De-nosing were carried out on some standard images to

the Matlab programs of the Slantlet transform. Some 2D inputs include images of a cameraman, MRI and Testpart. Implementation is carried out using Matlab.

7.1. Simulation Results of Compression

The Compression Ratio (CR) and Reconstruction Ratio (RR) for different input images are tabulated as shown in the table 6.1 and table 6.2. The graphs are plotted against various threshold values and the ratios.

Thr value	Cameraman		MRI		Testpart	
	% CR	% RR	% CR	% RR	% CR	% RR
0	0	100	0	100	0	100
0.5	14.73	100	52.9	99.9	11.6	100
0.05	64.68	99.9	70.1	99.9	54.2	99.99
0.05	19.0813	9	21.2111	7	19.0790	5
10.0	76.02	99.9	79.9	99.8	69.9	99.99
0.5	19.2087	9	21.0978	7	19.1190	2
5.0	20.0354	9	21.8754	7	16.8582	2
100	98.02	99.8	98.8	94.6	95.9	99.69
10.0	20.9197	9	22.7155	7	20.6429	5
20.0	22.4248	9	24.1644	7	21.9417	7
500.0	99.8123	99.20	99.825	99.20	99.622	99.007
5000	99.97	96.2	99.9	18.7	99.9	94.74

Table 7.1. Percentage of Compression Ratio(CR) and Reconstruction Ratio(RR) for different Images

Figure 7.1. Graph for Cameraman

7.2. Simulation results of de-noising

For the various threshold values, the percentage of signal to noise ratio for different images are tabulated as shown in the table 6.3 and 6.4. The graphs are plotted against various threshold values and the PSNR.

Table 7.2. Percentage of Signal to noise ratio (PSNR) for different Images

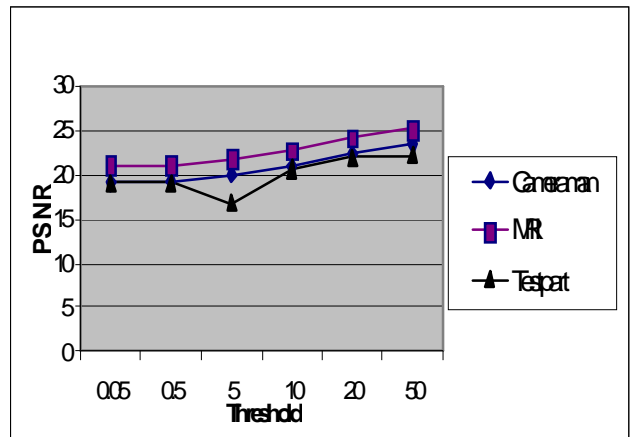
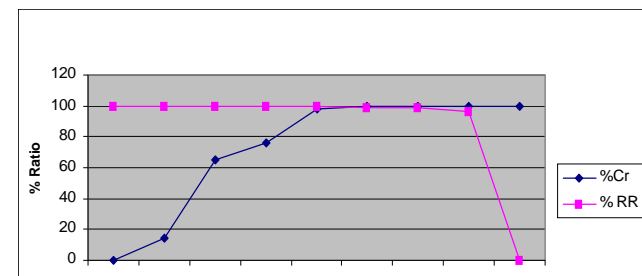


Figure 7.2. Graph between Threshold and PSNR for different Images.



8. Snapshots

The snapshots of the implementation of Compression & De-noising for different images are shown

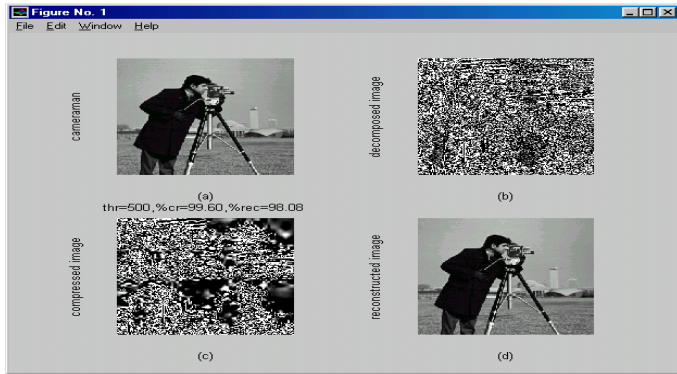


Figure 8.1. (a) Input image(cameraman) (b) Decomposed image (c) Compressed image (d) Reconstructed image.

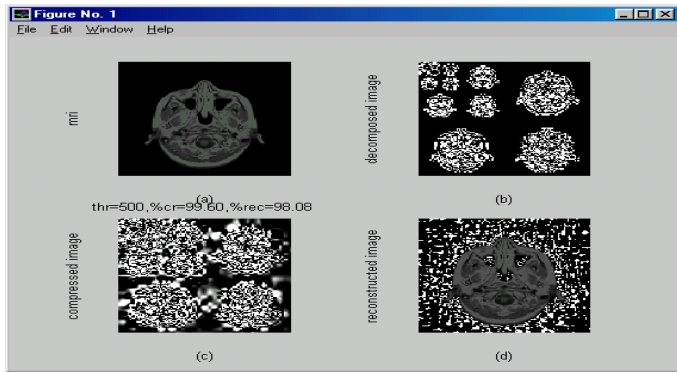


Figure 8.2. (a) Input image(MRI) (b) Decomposed image (c) Compressed image (d) Reconstructed image.

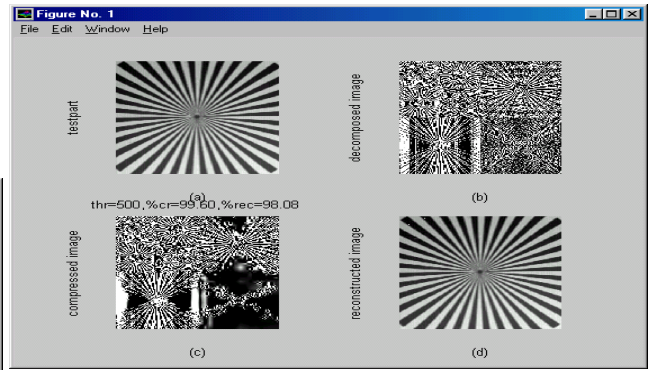


Figure 8.3. (a) Input image(testpart) (b) Decomposed image (c) Compressed image (d) Reconstructed image

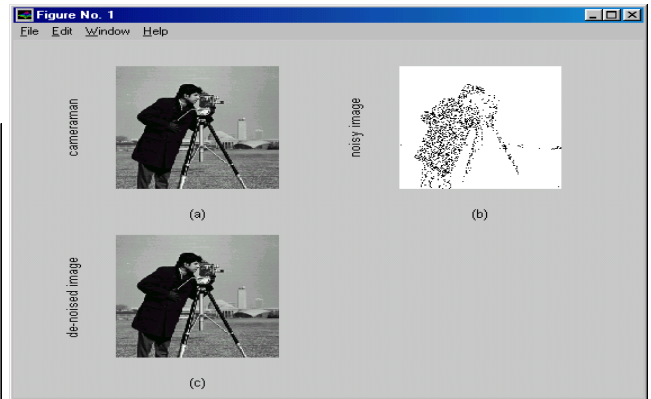


Figure 8.4. (a) Input image(cameraman) (b) Noisy image (c) De-noised image .

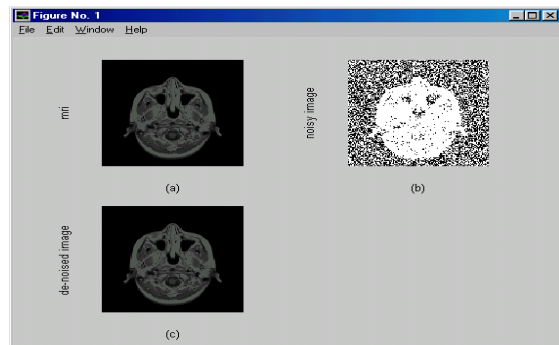


Figure 8.5. (a) Input image(MRI) (b) Noisy image (c) Denoised image.

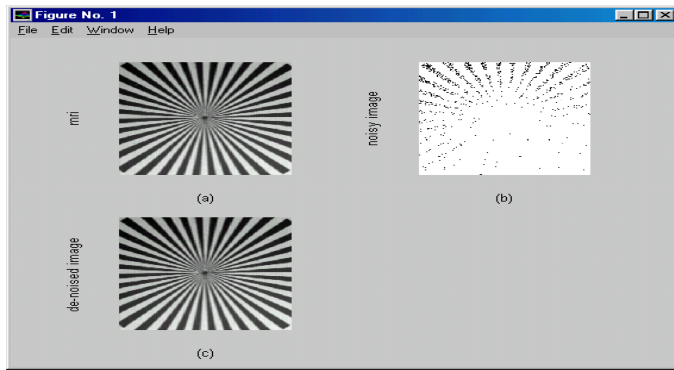


Figure 8.6. (a) Input image(Testpart) (b) Noisy image (c) De-noised image.

9. Conclusion

The Slantlet transform orthogonal filterbank uses the filters of shorter length as compared to iterated 2-scale DWT filterbank. The number of filters increases according to the order. The Slantlet transform gives better compression result for the piecewise linear data. The Slantlet transform has been applied for compression and de-noising of various images. The matlab programs for the Slantlet transform applied for compression and de-noising are developed and the computer simulation is carried out for some test images. It is observed that as the threshold level increases better compression ratio and PSNR can be achieved for the test data.

10. Future scope

The present work can be extended to compare the performance of applying Wavelet transform and Slantlet transform on images and also study the characteristics/performance benefits of Slantlet transform over Wavelet transform. This work can be enhanced to wavelet packets, which offers a more complex and flexible analysis, because the details as well as the approximations are split.

Acknowledgements

Authors thank Singhania University Rajasthan and Vidya Vikas Institute of Technology, Mysore.

Bibliography

1. I. W. Selesnick, "The Slantlet Transform," IEEE Transaction on signal processing, vol 47, No. 5, May 1999.
2. David I. Donoho "Smooth wavelet decompositions with blocky coefficient kernels " May, 1993.
3. Vaidyanathan, p. p., "Multirate systems and filter banks", prentice hall inc., Englewood cliffs, new jersey, 1993.
4. M. Lmg, H. Guo, et. al "Noise reduction using an undecimated discrete wavelet transform", IEEE signal processing , vol 3, no. 1, Jan 1996.
5. Aali M. Reza , "From Fourier transform to Wavelet transform", October 27, 1999 white paper.
6. Paolo Zatelli, Andrea Antonello, " New Grass modules for multiresolution analysis with wavelets", proceedings of the open source gis - grass users conference, September 2002.
7. Kenneth Lee, Sinhye Park, et.al, "Wavelet-based image and video compression", Tcom 502, April, 1997.
8. Dr. Guangyi Chen , "Signal or Image denoising using wavelets, multiwavelets and curvelets", 2004.

Separation of Concerns in VoiceXML Applications

Sukhada P. Bhingarkar

Abstract— Many commercial applications provide customer services over the web like flight tracking, emergency notification, order inquiry etc. VoiceXML is an enabling technology for creating streamlined speech-based interface for such web-based information services. Whereas in computing, aspect-oriented programming (AOP) is a programming paradigm, which aims to increase modularity. AOP includes programming methods and tools that support the modularization of concerns at the level of the source code. The aim of this paper is to integrate AOP with VoiceXML. Aspect-Oriented Programming (AOP) encapsulates common low-level scattered code within reusable components called aspects. There are certain tags in VoiceXML like '<nomatch>', '<noinput>', '<error>' which appear commonly in every VoiceXML document. These tags can be considered as the concerns and can be put inside an aspect. This eliminates the need to programmatically write these tags in every VoiceXML document and modularizes the crosscutting-concerns.

Index Terms— AOP, AspectJ, ASR, IVR, TTS, VoiceXML, VXML

1 INTRODUCTION

The WWW has become primary source of information all over the world and accounts for the major proportion of entire Internet traffic. The next leading edge for the research on the web is to make it accessible via voice and audio. Considerable work has been done in this direction, which includes the design of VoiceXML and voice browsers. It allows voice applications to be developed and deployed in an analogous way to HTML for visual applications.

Aspect orientation is not a completely new approach to writing software. Aspect orientation is becoming a commonly adopted and de facto approach to practicing older ideas that can be traced to almost the beginning of software development. Development environments and tools that weave code, pragma instructions, and even debuggers all contain some of the behavior that underlies the aspect-oriented approach. It is a more modular implementation of the advantages that these technologies have brought to their own domains in the past. Aspect-oriented programming entails breaking down program logic into distinct parts called as *concerns*. All programming paradigms support some level of grouping and encapsulation of concerns into separate, independent entities by providing abstractions (e.g., procedures, modules, classes, methods) that can be used for implementing, abstracting and composing these concerns. But some concerns defy these forms of implementation and are called *crosscutting concerns* because they "cut across" multiple abstractions in a program. This paper

tries to integrate this feature of AOP with VoiceXML. VoiceXML has certain tags which appear across the scope of every VoiceXML page. These tags are crosscutting concerns that can be defined in a modular fashion with the help of aspect-oriented programming. This helps modularization and code reuse in VoiceXML based applications.

The rest of the paper is organized in following way: section II presents related work in this field. Section III and IV discusses VoiceXML and AOP respectively. The paper, then, proposes an integration of AOP with VoiceXML in Section V. Finally, the paper finishes with conclusion in Section VI.

2 RELATED WORK

During past few years, several applications are developed to assist visually impaired, technologically uneducated and underprivileged people to acoustically access the information that was originally intended to be accessed visually via a personal computer (PC). Voice response facilities are used for various kinds of information over the phone: time, weather, horoscopes, sports, cultural events and so on. Nuance is one of the providers for speech and imaging solutions for businesses and customers around the world that is based on IVR systems. The VoiceXML along with Nuance 8.5 speech recognition software can reduce cost and effort of deploying voice-driven services. In [1], a VoiceXML-driven audio wiki application is presented that is

accessible via both the Public Switched Telephone Network and the Internet. Users may access wiki content via fixed or mobile phones or via a PC using web browser or a Voice over IP service. Silog is a biometric authentication system that extends the conventional PC logon process using voice verification [2]. SeeCCT is a prototype of a multimodal social networking system designed for sharing geographical bookmarks [3]. In [4], VoiceXML portal is developed that allows people with a mobile or usual phone to get informed about cultural activities by dialing a phone number and by interacting with a computer via voice. Domain-specific dialogs are created in native languages viz. slavic languages using VoiceXML [5]. HearSay is a non-visual Web browser developed for visually impaired users [6].

AspectJ is one of the oldest and well-known aspect languages and it helped to bring AOP to the mainstream. AspectJ is an extension of the Java language which defines a special syntax for declaring aspects. The first versions of AspectJ featured compile-time source code weaving and bytecode weaving. It later merged with AspectWerkz which brought load-time weaving as well as AspectWerkz's annotation style to the language. AOP is used in a variety of fields. Aspect-oriented software development had played an important role in the design and implementation of PUMA which is a framework for the development of applications that analyze and transform C or C++ source code [7]. A framework for middleware design is invented which is based on the Concurrent Event-based Aspect-Oriented paradigm [8]. A fully dynamic and reconfigurable monitoring system is designed based on the concept of Adaptable Aspect-Oriented Programming (AAOP) in which a set of AOP aspects is used to run an application in a manner specified by the adaptability strategy [9]. The model can be used to implement systems that are able to monitor an application and its execution environment and perform actions such as changing the current set of resource management constraints applied to an application if the application/environment conditions change. An aspect-oriented approach is advocated as an improvement to the object oriented approach in dealing with the issues of code tangling and scattering in case of multilevel security [10].

3 VOICEXML

VoiceXML is HTML of Voice Web. It is W3C's standard XML format for specifying interactive voice dialogues between a human and a computer. VoiceXML is a markup language for creating voice user interfaces. It uses

• *Sukhada Bhingarkar is with the Computer Engineering Dept., MIT college of Engineering, Kothrud, Pune, India – 411038*
E-mail:sukhada.bhingarkar@gmail.com

Automatic Speech Recognition (ASR) and/or touchtone (DTMF keypad) for input, and prerecorded audio and text-to-speech (TTS) synthesis for output. Numerous commercial vendors such as IBM, TellMe and BeVocal provide voice browsers that can be used to "play" VoiceXML documents. Current voice interfaces to the web are of two types: voice interface to screen display and voice-only interface. The dynamic voice interface presented in this paper is a voice-only interface that uses a telephone as an input and output device. Figure 1 shows the core architecture of VoiceXML applications.

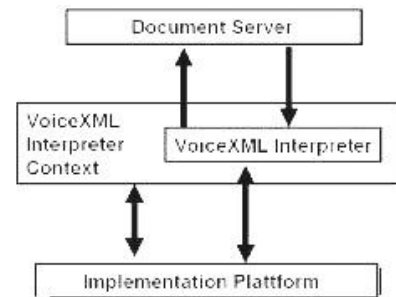


Fig. 1. Architectural Model of VoiceXML

A document server processes requests from a client application, the VoiceXML Interpreter, through the VoiceXML interpreter context. The server produces VoiceXML documents in reply, which are processed by the VoiceXML Interpreter. The VoiceXML interpreter context may monitor user inputs in parallel with the VoiceXML interpreter. The implementation platform is controlled by the VoiceXML interpreter context and by the VoiceXML interpreter.

4 ASPECT ORIENTED PROGRAMMING (AOP)

The main idea of AOP is to isolate the cross-cutting concerns from the application code thereby modularizing them as a different entity. A cross-cutting concern is behavior, and often data, that is used across the scope of a piece of software. It may be a constraint that is a characteristic of your software or simply behavior that every class must perform. The most common example of a cross-cutting concern is that of logging. Logging is a cross-cutting concern because it affects many areas across the software system and it intrudes on the business logic. Logging is potentially applied across many classes, and it is this form of horizontal application of the logging aspect that gives cross-cutting its name. Some central AOP concepts are as follows:

Aspects: A modularization of a concern that cuts across multiple objects. Transaction management is a good example of a crosscutting concern in J2EE applications.

Join Point: A point during the execution of a program, such as the execution of a method or the handling of an exception

Advice: The code that is executed when an aspect is invoked is called *advice*. Advice contains its own set of rules as to when it is to be invoked in relation to the join point that has been triggered. Different types of advice include "around," "before" and "after" advice.

Pointcut: A predicate that matches join points. Advice is associated with a pointcut expression and runs at any join point matched by the pointcut (for example, the execution of a method with a certain name).

Figure 2 shows the relationships between join points, aspects, pointcuts, advice, and your application classes.

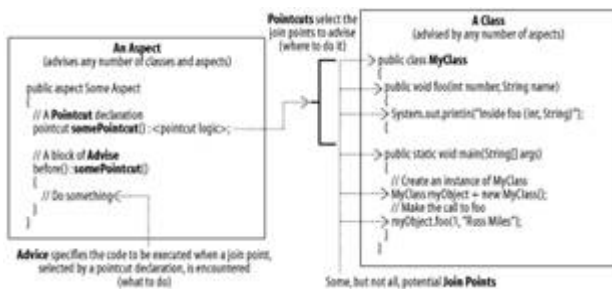


Fig.2. Relationship of AOP concepts with application class [11]

Figure 3 shows a simple business logic class along with an aspect applied to this class.

```
// Business logic class
package com.aop;
public class MyClass
{
    public void foo(int number, String name)
    {
        System.out.println("Inside foo (" + number + ", " + name + ")");
    }

    public static void main(String[] args)
    {
        MyClass myObject = new MyClass();
        myObject.foo(1, "Ram Sham");
    }
}

// An Aspect applied to the class
public aspect HelloWorld
{
    pointcut callPointcut() : call(void com.aop.MyClass.foo(int, String));

    before(): callPointcut()
    {
        System.out.println("Hello World");
        System.out.println("In the advice attached to the call pointcut");
    }
}
```

Fig.3. Business logic class along with aspect

5 INTEGRATING AOP WITH VOICEXML

VoiceXML tags such as '<nomatch>', '<noinput>' '<error>' are usually common tags in most of the VoiceXML documents. Figure 4 shows VoiceXML example that illustrates basic capabilities of VoiceXML. This VoiceXML is logically divided into three parts: header, body and footer. The header part can contain tags showing xml and vxml version. The body part contains core business logic while footer part has error handling and <nomatch>, <noinput> tags.

Fig. 4. VoiceXML example

The commonly occurring tags in VoiceXML can be considered as crosscutting concerns and are encapsulated in a special class i.e. an aspect. Figure 5 shows the Servlet that is used to construct VoiceXML.

```
public class AOServlet extends HttpServlet
{
    public void doGet(HttpServletRequest request, HttpServletResponse response) throws ServletException, IOException
    {
        ServletOutputStream out = response.getOutputStream();
        // core business logic code
    }
}
```

Fig. 5. Servlet generating VoiceXML code

doGet() or doPost() methods of Servlet concentrates only on core logic code i.e. body part of VoiceXML. The header and footer parts of VoiceXML will be taken care by an aspect class shown in figure 6.

```

public aspect AddXMLHeaderAndFooter
{
    // Define pointcut for doGet(HttpServletRequest, HttpServletResponse)
    before (HttpServletRequest request, HttpServletResponse response) throws
    IOException: captureHttpRequest (request, response)
    {
        ServletOutputStream out=response.getOutputStream();
        out.println("<?xml version='1.0' encoding='iso-8859-1'?>");
        out.println("<?xml version='2.1'?>");
    }
    after (HttpServletRequest request, HttpServletResponse response) throws
    IOException: captureHttpRequest (request, response)
    {
        ServletOutputStream out=response.getOutputStream();
        out.println("<error>");
        out.println("<if cond='$_event == 'error.badfetch'?'>");
        out.println("<log expr='$_' *** ERROR.BADFETCH CACHED*** '\>");
        out.println("</if>");
        out.println("</error>");
    }
    // Rest of code handles nomatch, noinput etc. tags.
}

```

Fig. 6. Aspect applied to Servlet

6 CONCLUSION

VoiceXML is a language to create voice-user interfaces while AOP allows us to dynamically modify our static model to include the code required to fulfill the secondary requirements without having to modify the original static model. Integration of AOP with VoiceXML better separates the concerns of Voice based applications, thereby providing modularization. This helps developers to concentrate on core logic and promotes code reuse.

REFERENCES

- [1] Constantinos Koliass et al, "Design and implementation of a VoiceXML-driven wiki application for assistive environments on the web", Personal and Ubiquitous Computing, 2010, Volume 14, Number 6, 527-539, Springer-Verlag London Limited 2010
- [2] Sergio Grau, Tony Allen, Nasser Sherkat, "Silog: Speech input logon", Knowledge-Based Systems, Volume 22, Issue 7, October 2009, Pages 535-539
- [3] Stan Kurkovsky et al., "Mobile Voice Access in Social Networking Systems", in the Proceedings of 5th IEEE International Conference on Information Technology: New Generations, Las Vegas, USA, 7-9 Apr. 2008
- [4] Evangelia Boufardea et al, "A Dynamic Voice Portal for Delivery of Cultural Content" in the Proceedings of 3rd International Conference on Internet and Web Applications and Services (ICIW'08), Athens, 8-13 Jun. 2008
- [5] Brkic M., Matetic M., "VoiceXML for Slavic Languages Application Development", in the Proceedings of IEEE international conference on Human System Interactions, Krakow, Poland, 25-27 May 2008
- [6] Borodin Y, Mahmud J, Ramakrishnan IV, Stent A (2007) The hearsay non-visual web browser. In: ACM international conference proceeding series, proceedings of the 2007 international cross-disciplinary conference on web accessibility (W4A), vol 225. Banff, Canada, pp 128-129
- [7] Matthias Urban, Daniel Lohmann, Olaf Spinczyk, "The Aspect-Oriented Design of the PUMA C/C++ Parser Framework". In: AOSD '10: Proceedings of the 9th International Conference on

Aspect-Oriented Software Development, March 2010.

- [8] Edgar Marques, Luís Veiga, Paulo Ferreira, "An extensible framework for middleware design based on concurrent event-based AOP", ARM '10: Proceedings of the 9th International Workshop on Adaptive and Reflective Middleware, November 2010
- [9] Arkadiusz Janik, Krzysztof Zielinski, "AAOP-based dynamically reconfigurable monitoring system", Information and Software Technology, Volume 52 Issue 4, April 2010
- [10] S. Kotrappa, Prakash J. Kulkarni, "Multilevel Security Using Aspect Oriented Programming AspectJ", ARTCOM '10: Proceedings of the 2010 International Conference on Advances in Recent Technologies in Communication and Computing, October 2010
- [11] Russell Miles, "AspectJ Cookbook", O'Reilly, 2004

Error of Approximation in Case of Definite Integrals

Rajesh Kumar Sinha, Satya Narayan Mahto, Dhananjay Sharan

Abstract— This paper proposes a method for computation of error of approximation involved in case of evaluation of integrals of single variable. The error associated with a quadrature rule provides information with a difference of approximation. In numerical integration, approximation of any function is done through polynomial of suitable degree but shows a difference in their integration integrated over some interval. Such difference is error of approximation. Sometime, it is difficult to evaluate the integral by analytical methods Numerical Integration or Numerical Quadrature can be an alternative approach to solve such problems. As in other numerical techniques, it often results in approximate solution. The Integration can be performed on a continuous function on set of data.

Index Terms— Quadrature rule, Simpsons rule, Chebyshev polynomials, approximation, interpolation, error.

1 INTRODUCTION

TO evaluate the definite integral of a function that has no explicit antiderivative of whose antiderivative is not easy to obtain; the basic method involved in approximating is numerical quadrature [1]-[4].

$$\int_a^b f(x)dx \quad (1)$$

i.e. $\sum_{i=0}^n \alpha_i f(x)$ to approximate $\int_a^b f(x)dx$. The methodology for computing the antiderivative at a given point, the polynomial $p(x)$ approximating the function $f(x)$ generally oscillates about the function. This means that if $y = p(x)$ over estimates the function $y = f(x)$ in one interval then it would underestimate it in the next interval [5]. As a result, while the area is overestimated in one interval, it may be underestimated in the next interval so that the overall effect of error in the two intervals will be equal to the sum of their moduli, instead the effect of the error in one interval will be neutralized to some extent by the error in the next interval. Therefore, the estimated error in an integration formula may be unrealistically too high. In view to above discussed facts, the paper would reveal types of approximation following the condition 'best' approxima-

tion for a given function, concentrating mainly on polynomial approximation. For approximation, there is considered a polynomial of first degree such as $y = a + bx$ a good approximation to a given function for the interval (a, b) .

2 PROPOSED METHOD

2.1 Reflection on Approximation

This section cover types of approximation following the condition 'best' approximation for a given function, concentrating mainly on polynomial approximation. In this for approximation, there is considered a polynomial of first degree such as $y = a + bx$; a good approximation to a given continuous function for the interval $(0, 1)$.

Under the assumption of given concept two following statements may be considered as,

The Taylor polynomial at $x = 0$ (assuming $f'(0)$ exists)

$$y = f(0) + xf'(0) \quad (2)$$

The interpolating polynomial constructed at $x = 0$ and $x = 1$.

$$y = f(0) + x[f(1) - f(0)] \quad (3)$$

A justification may be laid that a Taylor or interpolating polynomial constructed at some other point would be more suitable. However, these approximations are designed to initiate the behavior of f at only one or two points.

- Rajesh Kumar Sinha is with the Department of Mathematics, NIT Patna, India. E-mail: rajesh_nitpat@rediffmail.com
- Satya Narayan Mahto is with the Department of Mathematics, M. G. College, LNMU, Darbhanga, India.
- Dhananjay Sharan is research scholar in the Department of Mathematics, NIT Patna, India.

Since, the polynomial of first degree in x as shown above $y = a + bx$ follows a good approximation to f throughout the interval $(0,1)$. Now, for values of a and b , the required mathematical exists such as $\max_{0 \leq x \leq 1} [f(x) - (a + bx)]$ is minimized over all choices of two values a and b . This expression is said as minimax (or Chebyshev) approximation. Instead of minimizing the minimum error between the (continuous) function f and the approximating straight-line, the process of maximizing 'sum' of the moduli of the errors may be undertaken.

For values of a and b , $\int_0^1 [f(x) - (a + bx)] dx$ is minimized that is called a base L_1 approximation. It should be noted that the L_1 approximation provides equal weight to all the errors, while the minimax approximation approximate in the error of largest modulus. Again, stressing on other approximation f which in a sense, lies between the extremes of L_1 and minimax approximation. Also, for a fixed value of $p \geq 1$, two values a and b are formed so that $\int_0^1 [f(x) - (a + bx)]^p dx$ is minimized and therefore would be suggested as best L_p approximation. The above maximized expression followed that due to the presence of the P^{th} power, the error of largest modulus tends to dominate as p increases with f continuous. It can be shown that, as $p \rightarrow \infty$ the best L_p approximation tends to the minimax approximation which is therefore sometimes called the best L_∞ approximation. Thus the L_p approximations consist of a spectrum ranging from the L_1 to the minimax approximations. Further, for $1 < p < \infty$, L_2 approximation, is the only commonly used and is better known as the best square approximation.

2.2 Generalized case for approximation under certain interval

This section would reveal for giving light on consideration of methods of approximating to f' , given the values of $f(x)$ at certain points. If p is same polynomial, approximation to f an application of p' would exist an approximation to f' . However, there is need to be careful; the maximum modulus of $f'(x) - p'(x)$ on a given interval (a, b) can be much larger than the maximum modulus of $f(x) - p(x)$.

To make an evident proof for a given statement an assumption is made

$$f(x) - p(x) = 10^{-2} T_n(x) \tag{4}$$

Where $T_n(x)$ is a polynomial of degree n in x with leading term $2^{n-1} x^n$ for $n > 0$. The T_n are known as Chebyshev polynomials, the Russian mathematician P. L. Chebyshev (1821 -1894) who has contributed the equation (4) for the difference between of the first approximation and the polynomial. Now, to determine the turning value, of T_n the derivative

$$\frac{d}{dx} T_n(x) = \frac{d}{d\theta} \cos n\theta \frac{d\theta}{dx} = \frac{n \sin n\theta}{\sin \theta} \tag{5}$$

Since $\frac{d\theta}{dx} = \frac{1}{\frac{dx}{d\theta}} = -\frac{1}{\sin \theta}$ (6)

i.e. $T'_n(x) = \frac{n \sin n\theta}{\sin \theta} = n^2 \left(\frac{\sin n\theta}{n\theta} \right) \left(\frac{\theta}{\sin \theta} \right)$ (7)

but $\lim_{\theta \rightarrow 0} \left(\frac{\sin \theta}{\theta} \right) = 1$ (8)

Thus $T'_n(1) = n^2$ (9)

It can be shown that this is the maximum modulus of T'_n on $(-1, 1)$. If $n = 10$, say, the maximum modulus of $f - p = 10^{-2}$ on $(-1, 1)$ whereas $f' - p' = 1$. Furthermore, the consideration of an approximation f and the polynomial p_n that interpolates the approximation f at distinct point x_1, \dots, x_n . If such happens then there exists a number ξ_x (depending on x) in certain interval (a, b) such that

$$f(x) - p_n(x) = (x - x_0)(x - x_1) \dots (x - x_n) \frac{f^{(n+1)}(\xi_x)}{(n+1)!} \tag{10}$$

$$f'(x) - p'_n(x) = \pi_{n+1}(x) \frac{f^{(n+1)}(\xi_x)}{(n+1)!} \tag{11}$$

where $\pi_{n+1}(x) = (x - x_0)(x - x_1) \dots (x - x_n)$.

Thus equation (11) is known as error that exists in following statement. If (a, b) is any interval and also contains (n + 1) points x, x_0, x_1, \dots, x_n . Suppose further it is assumed that $f, f', \dots, f^{(n)}$ exist and are continuous on the interval (a, b) and $f^{(n+1)}$ exists for $a < x < b$ then the error holds as shown in equation (11). Thus, $f^{(n+1)}$ exists on some interval (a, b) which is $x, x_0, x_1, x_2, \dots, x_n$. Here, the number $\xi_x \in (a, b)$ (depending on x). Differentiating (11) with respect to x,

$$f'(x) - p'_n(x) = \pi'_{n+1}(x) \frac{f^{(n+1)}(x)}{(n+1)} + \frac{\pi_{n+1}(x)}{(n+1)} \frac{d}{dx} f^{(n+1)}(\xi_x) \quad (12)$$

In general, there is nothing further to state about the second term on the right of equation (12). We can not perform the differentiation with respect to x of $f^{(n+1)}(\xi_x)$, since ξ_x is an unknown function of x. Thus, for Integral values of x given in equation (12) is unless for determining the accuracy with which p'_n approximately to f' . However, if we restrict x to one of the values x_0, x_1, \dots, x_n , then $\pi_{n+1}(x) = 0$ and the unknown second term on the right of equation (12) becomes zero.

$$f'(x_r) - p'_n(x) = \pi'_{n+1}(x_r) \frac{f^{(n+1)}(\xi_r)}{(n+1)} \quad (13)$$

Where ξ_r has been considered when $x = x_r$. Now, applying forward difference to express the polynomial $p_n(x)$ in given form such that

$$p_n(x) = p_n(x_0 + sh) \quad (14)$$

$$p_n(x) = f_0 + \binom{s}{1} \Delta f_0 + \binom{s}{2} \Delta^2 f_0 + \dots + \binom{s}{n} \Delta^n f_0 \quad (15)$$

$$\text{Since } x = x_0 + sh \quad (16)$$

$$\text{i.e. } \frac{dx}{ds} = h \quad (17)$$

Also

$$p_n(x) = p_n(x_0 + sh) \quad (18)$$

Differentiating both sides with respect to x ,

$$p'_n(x) = \frac{ds}{dx} \frac{d}{ds} p_n(x_0 + sh) \quad (19)$$

$$p'_n(x) = \frac{1}{dx} \left[\frac{d}{ds} \left\{ f_0 + \binom{s}{1} \Delta f_0 + \binom{s}{2} \Delta^2 f_0 + \dots + \binom{s}{n} \Delta^n f_0 \right\} \right] \quad (20)$$

$$p'_n(x) = \frac{1}{h} \left[\Delta f_0 + \frac{1}{2} (2s-1) \Delta^2 f_0 + \dots + \frac{d}{ds} \binom{s}{n} \Delta^n f_0 \right] \quad (21)$$

To calculate π'_{n+1} ,

$$\pi_{n+1}(x) = (x - x_r) \prod_{j \neq r} (x - x_j) \quad (22)$$

where π'_{n+1} would be obtained by means of differentiating equations (22) such that

$$\pi'_{n+1}(x) = \left(\frac{d}{dx} (x - x_r) \right) \prod_{j \neq r} (x - x_j) + (x - x_r) \frac{d}{dx} \prod_{j \neq r} (x - x_j) \quad (23)$$

By putting $x = x_r$, the second term on the right of equation (22) becomes zero.

$$\text{Thus } \pi'_{n+1}(x) = \prod_{j \neq r} (x - x_j) = (-1)^{n-r} h^n r! (n-r)! \quad (24)$$

Since

$$x_r - x_j = (r - j)h \quad (25)$$

From (13)

$$f'(x_r) - p'_n(x_r) = (-1)^{(n-r)} h^n \frac{r!(n-r)!}{(n+r)!} f^{(n+1)}(\xi_r) \quad (26)$$

Furthermore, investigating the case of polynomial, an interpretation hold that polynomials are sufficiently accurate for many approximation and interpolation tasks.

2.3 Degree of Accuracy

Now the degree of accuracy of quadrature formula is the largest positive integer n such that the formula is exact for x^k , for each $k = 0, 1, \dots, n$. The Trapezoidal and Simpson's rule have degrees of precision one and three, respectively. Integration and summation are linear operations; that is,

$$\int_a^b (\alpha f(x) + \beta g(x)) dx = \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx \quad (27)$$

$$\sum_{i=0}^n (\alpha f(x_i) + \beta g(x_i)) = \alpha \sum_{i=0}^n f(x_i) + \beta \sum_{i=0}^n g(x_i) \quad (28)$$

For each pair of integral functions f and g and each pair of real constants. This implies that the degree of precision of a quadrature formula in n if and only if the error $E(p(x)) = 0$ for all polynomials $p(x)$ of degree $k = 0, 1, \dots, n$, but $E(p(x)) \neq 0$ for some polynomial $p(x)$ of degree $(n+1)$.

3 CONCLUSION

Increasing, the degree of the approximating polynomial does not guarantee better accuracy. In a higher degree polynomial, the coefficients also get bigger which may magnify the errors. Similarly, reducing the size of the sub-interval by increasing their number may also lead to accumulation of rounding errors. Therefore a balance should be kept between the two, i.e. degree of polynomial and total number of intervals. These are the primary motivations for studying the techniques of numerical integration/quadrature [6]-[9]. In case of Simpson's rule technique individually to the subintervals $[a, (a+b)/2]$ and $[(a+b)/2, b]$; use error estimation procedure to determine if the approximation to the integral on subinterval is within a tolerance of $\varepsilon/2$. If so, then sum the approximations to procedure an approximation of function $f(x)$ over interval (a, b) within the tolerance ε . If the approximation on one of the subintervals fails to be within the tolerance $\varepsilon/2$, then that subinterval is itself subdivided, and the procedure is reapplied to the two subintervals to determine if the approximation on each subinterval is accurate to within $\varepsilon/4$. This halving procedure is continued until each portion is within the required tolerance. Thus, Numerical analysis is the study of algorithms that use numerical approximation for the problems of continuous functions [10]-[12]. Numerical analysis continues this long tradition of practical mathematical calculations. Much like the Babylonian approximation, modern numerical analysis does not seek exact answers, since the exact answers are often impossible to obtain in practice. Instead, much of numerical analysis is concerned with obtaining approximate solutions while maintaining reasonable bounds on errors. It finds applications in all fields of engineering and the

physical sciences.

REFERENCES

- [1] K. E. Atkinson, An Introduction to Numerical Analysis, Wiley, New York, 1993.
- [2] R. E. Beard, "Some notes on approximate product integration," *J. Inst. Actur.*, vol. 73, pp. 356-416, 1947.
- [3] C. T. H. Baker, "On the nature of certain quadrature formulas and their errors," *SIAM. J. Numer. anal.*, vol. 5, pp. 783-804, 1968.
- [4] P. J. Daniell, "Remainders in interpolation and quadrature formulae," *Math. Gaz.*, Vol. 24, pp. 238-244, 1940.
- [5] R. K. Sinha, "Estimating error involved in case of Evaluation of Integrals of single variable," *Int J. Comp. Tech. Appl.*, Vol. 2, No. 2, pp. 345-348, 2011.
- [6] T. J. Akai, Applied Numerical Methods for Engineers, Wiley, New York, 1993.
- [7] L. M. Delves, "The Numerical Evaluation of Principal Value Integrals," *Computer Journal*, Vol. 10, pp. 389, 1968.
- [8] Brain Bradi, A Friendly Introduction to Numerical Analysis, pp. 441-532, Pearson Education, 2009.
- [9] R. K. Sinha, "Numerical Method for evaluating the Integrable function on a finite interval," *Int. J. of Engineering Science and Technology*, Vol. 2, No. 6, pp. 2200-2206, 2010.
- [10] C. E. Froberg, Introduction to Numerical Analysis, Addison-Wesley Pub. Co. Inc.
- [11] Ibid, The Numerical Evaluation of class Integrals, Proc. Comb. Phil. Soc. 52.
- [12] P. J. Davis and P. Rebinowitz, Method of Numerical Integration, 2nd edition, Academic Press, New York, 1984.

An Assessment model for Intelligence Competencies of Accounting Information Systems

Mehdi Ghazanfari, Mostafa Jafari, Saeed Rouhani

Abstract— Accounting Information Systems (AIS) as computer-based systems that processes financial information and supports decision tasks have been implemented in most organizations but, but they still encounter a lack of Intelligence in their decision-making processes. Models and methods to evaluate and assess the Intelligence-level of Accounting Information Systems can be useful in deploying suitable business intelligence (BI) services. This paper discusses BI Assessment criteria, fundamental structure and factors used in the Assessment model. Factors and the proposed model can assess the intelligence of Accounting Information Systems to achieve enhanced decision support in organizations. The statistical analysis identified five factors of the Assessment model. This model helps organizations to design, buy and implement Accounting Information Systems for better decision support. The study also provides criteria to help organizations and software vendors implement AIS from decision support perspectives.

Index Terms— Business Intelligence; Decision Support; Accounting Information Systems; Assessment Model.

1 INTRODUCTION

Information and knowledge represent the fundamental wealth of an organization. Enterprises try to utilize this wealth to gain competitive advantage when making important decisions. Enterprise systems like Accounting Information Systems (AIS) converts and store the data. Therefore, it is important to integrate decision-support into the environment of these systems. Business intelligence (BI) can be embedded in these enterprise systems to obtain this competitive advantage.

Today, approaches using an individual system for decision-support, such as decision-support systems (DSS), have been replaced by a new, environmental approach. In the past, DSS were independent, separate systems in an organization (island systems). However, enterprise systems are now the foundation of an organization, and practitioners are designing BI as an umbrella concept to create a decision-support environment for management (Alter 2004). The increasing trend to use intelligent tools in business systems has increased the need for Intelligence Assessment of Accounting Information Systems (AIS).

There have been some limited efforts to assess BI, but they have always considered BI as a system that is isolated from the enterprise systems like AIS. Taking a global view, Lönnqvist and Pirttimäki (2006) have designed BI

performance measures, but before their effort, the measurement and the evaluation in the BI field were restricted to proving the worth of BI investment, and the value of BI. Elbashir et al. (2008) have discussed measuring the effects of BI systems on the business process, and have presented effective methods of measurement. Lin et al. (2009) have also developed a performance evaluation model for BI systems using ANP, but they have also treated BI as a separate system.

ANP, but they have also treated BI as a separate system.

Organizations usually utilize functional and non-functional requirements to assess and select enterprise systems like AIS, so the consideration of their decision-support environment as a non-functional requirement, raises the following questions.

1. Which criteria are suitable and effective in the Intelligence assessment of Accounting Information Systems (AIS)?
2. What is the fundamental structure of these criteria?

This research was carried out to find answers to the above questions and to provide a model for efficient decision-support by evaluating systems and making BI an integral part of these systems. The rest of this paper is organized as follows. Section 2, describes brief literature about AIS. Section 3 is about attempts in previous studies to define Business intelligence (BI). A wide-ranging literature review about BI and decision-support criteria to assess systems is also summarized in Section 3. Research methodology and research stages are discussed in section 4. Section 5 discusses the design of the questionnaire, data collection, reliability analysis, factor extraction, and label-

- Mehdi Ghazanfari is currently full professor in Iran University of Science and Technology, Tehran, Iran, E-mail: Mehdi@iust.ac.ir
- Mostafa Jafari is currently assistant professor in Iran University of Science and Technology, Tehran, Iran
- Saeed Rouhani is currently PhD Candidate in Iran University of Science and Technology, Tehran, Iran, E-mail: SRouhani@iust.ac.ir, Contact: +989122034980

ling and assessment model. Finally, Section 6 concludes the research work, its findings and proposed future research.

2 ACCOUNTING INFORMATION SYSTEMS

An Accounting Information Systems (AIS) is defined as a computer-based system that processes financial information and supports decision tasks in the context of coordination and control of organizational activities. Extant accounting information systems research has evolved from the source disciplines of Computer science, organizational theory and cognitive psychology. An advantage of this evolution is a diverse and rich literature with the potential for exploring many different interrelationships among technical, organizational and individual aspects of judgment and decision performance. AIS research also spans from the macro to the micro aspects of the information system (Birnberg & Shields, 1989; Gelinias et al., 2005).

The comparative advantage of accounting researchers within the study of IT lies in their institutional accounting knowledge. Systems researchers can contribute insights into the development of systems utilizing technology, and the other sub-areas can contribute insights into the task characteristics in the environment. For instance, systems researchers have extensively investigated group decision support systems (GDSS), but they have only recently been considered in auditing. On the other hand, auditing research has extensively investigated the role of knowledge and expertise. The merging of the two sets of findings may be relevant to AIS design, training, and use.

As comparable term, Management accounting systems (MAS) also are formal systems that provide information from the internal and external environment to managers (Bouwens & Abernethy, 2000). They include reports, performance measurement systems, computerized information systems, such as executive information systems or management information systems, and also planning, budgeting, and forecasting processes required to prepare and review management accounting information.

Research on management accounting and integrated information systems (IIS) has evolved across a number of different research streams. Some research streams put heavier emphasis on the management accounting side, while other research streams put emphasis on the information systems side. Likewise, different research streams approach the topic from different perspectives (Anders Rom & Carsten Rohde, 2007).

A major stream of research within AIS research deals with the modeling of accounting information systems. Several modeling techniques stay alive within the information systems literature (e.g. entity-relationship diagrams, flowcharts and data flow diagrams). Whereas these modeling techniques can be used when modeling accounting information systems (Gelinias et al., 2005), But, the REA modeling technique is particular to the AIS domain. The REA model, which maps resources, events and

agents, was first described by McCarthy (1979, 1982) and later developed by an exclusive group of researchers (David et al., 2002). Extensions to resources, events and agents include locations (Denna et al., 1993), tasks and commitments (Geerts and McCarthy, 2002).

An unshakable stream of research exists within the AIS literature that investigates behavioral issues in relation to accounting information systems (Sutton and Arnold, 2002). This stream of research investigates the impact of IT on individuals, organizations and society.

An example of behavioral AIS research is a study carried out by Arnold et al. (2004) on the use and effect of intelligent decision aids. The authors find that smart machines must be operated by smart people. If users are inexperienced, they will be negatively impacted by the system. Furthermore, they will not learn by experience. Abernethy and Vagnoni (2004) found that top management uses the newly implemented system for monitoring. Use of AIS is found to have a positive effect on cost consciousness, but the cost consciousness is hampered if people have informal power. In this context, power is an explanatory variable of AIS use.

3 BUSINESS INTELLIGENCE

Business Intelligence or BI is a grand, umbrella term introduced by Howard Dresner of the Gartner Group in 1989 to describe a set of concepts and methods to improve business decision-making by using fact-based computerized support systems (Nylund, 1999). The first scientific definition, by Ghoshal and Kim (1986) referred BI to a management philosophy and tool that helps organizations to manage and refine business information for the purpose of making effective decisions.

BI was considered to be an instrument of analysis, providing automated decision-making about business conditions, sales, customer demand, and product preference and so on. It uses huge-database (data-warehouse) analysis, as well as mathematical, statistical, artificial intelligence, data mining and on-line analysis processing (OLAP) (Berson and Smith, 1997). Eckerson (2005) understood that BI must be able to provide the following tools: production reporting tools, end-user query and reporting tools, OLAP, dashboard/screen tools, data mining tools and planning and modelling tools.

BI includes a set of concepts, methods, and processes to improve business decisions, which use information from multiple sources and apply past experience to develop an exact understanding of business dynamics (Maria, 2005). It integrates the analysis of data with decision-analysis tools to provide the right information to the right persons throughout the organization with the purpose of improving strategic and tactical decisions. A BI system is a data-driven DSS that primarily supports the querying of an historical database and the production of periodic summary reports (Power, 2008).

Lönnqvist and Pirttimäki (2006), stated that the term,

BI, can be used when referring to the following concepts:

1. Related information and knowledge of the organization, which describe the business environment, the organization itself, the conditions of the market, customers and competitors, and economic issues;

2. Systemic and systematic processes by which organizations obtain, analyse, and distribute the information for making decisions about business operations.

A literature review around the theme of BI shows "division" between technical and managerial view points, tracing two broad patterns. The managerial approach sees BI as a process in which data, gathered from inside and outside the enterprise and are integrated in order to generate information relevant to the decision-making process. The role of BI here is to create an informational environment in which operational data gathered from transactional processing systems (TPS) and external sources can be analysed, in order to extract "strategic" business knowledge to support the unstructured decisions of management.

The technical approach considers BI as a set of tools that support the process described above. The focus is not on the process itself, but on the technologies, algorithms and tools that allow the saving, recovery, manipulation and analysis of data and information (Petrini and Pozzebon, 2008).

However, in the overall view, there are two important issues. First, the core of BI is the gathering, analysis and distribution of information. Second, the objective of BI is to support the strategic decision-making process.

By strategic decisions, we mean decisions related to implementation and Assessment of organizational vision, mission, goals, and objectives, which are supposed to have medium- to long-term impact on the organization, as opposed to operational decisions, which are day-to-day in nature and more related to execution (Petrini and Pozzebon, 2008).

4 RESEARCH METHODOLOGY

Based on literature review, the points discussed above, the authors' recent researches on BI and applying some statistical methods, the research structure of this study has been developed on seven stages as shown in Figure 1. In this way, at first stage a literature review was done on business intelligence specifications or criteria that a system should have to cover BI definitions. These criteria listed in Table 1.

At the Second stage, a questionnaire was designed with two main parts: first section of the questionnaire consisted of some questions the characteristics of the interviewees. The content of second section was based on business intelligence specifications which were asked as the important evolution criteria.

At the third and fourth stage, the survey is run to collect data from interviewees; and based on the collected data; the reliability analysis can be performed. Reliability anal-

TABLE 1
BI ASSESSMENT CRITERIA

Criteria ID	Criteria Name	Related Studies
C1	Group sorting tools & methodology (Groupware)	Shin et al. (2002), Reich and Kapeluk (2005), Damart et al. (2007), Marinoni et al. (2009)
C2	Group decision-making	Eom (1999), Evers (2008), Yu et al. (2009)
C3	Flexible models	Reich and Kapeluk (2005), Zack (2007), Lin et al. (2009)
C4	Problem clustering	Reich and Kapeluk (2005), Loebbecke and Huyskens (2007), Lamptey et al. (2008)
C5	Optimization technique	Lee and Park (2005), Nie et al. (2008), Shang et al. (2008), Azadivar et al. (2009), Delorme et al. (2009)
C6	Learning technique	Power and Sharda (2007), Ranjan (2008), Li et al. (2009), Zhan et al. (2009)
C7	Import data from other systems	Ozbayrak and Bell (2003), Alter (2004), Shang et al. (2008), Quinn (2009)
C8	Export reports to other systems	Ozbayrak and Bell (2003), Shi et al. (2007), Shang et al. (2008)
C9	Simulation models	Power and Sharda (2007), Shang et al. (2008), Quinn (2009), Zhan et al. (2009)
C10	Risk simulation	Evers (2008), Galasso and Thierry (2008)
C11	Financial analyses tools	Santhanam and Guimaraes (1995), Raggad (1997), Gao and Xu (2009)
C12	Visual graphs	Noori and Salmi (2005), Kwon et al. (2007), Power and Sharda (2007), Li et al. (2008), Azadivar et al. (2009)
C13	Summarization	Bolloju et al. (2002), Hemsley-Brown (2005), Power and Sharda (2007), Power (2008)
C14	Evolutionary prototyping model	Fazlabadi and Valadov (2001), Bolloju et al. (2002), Gao and Xu (2009), Zhang et al. (2009)
C15	Dynamic model prototyping	Koutsoukis et al. (2000), Bolloju et al. (2002), Goul and Corral (2007), Gonzalez et al. (2008), Fitz et al. (2008)
C16	Backward & forward reasoning	Gottschalk (2006), Evers (2008), Zhang et al. (2009)
C17	Knowledge reasoning	Ozbayrak and Bell (2003), Plessis and Toit (2006), Evers (2008)
C18	Alarms & warnings	Power (2008), Ross et al. (2008), Zhang et al. (2009)
C19	Dashboard/Recommender	Nemati et al. (2002), Hedgebeth (2007), Bose (2009)
C20	Combination of experiments	Courtney (2001), Nemati et al. (2002), Gottschalk (2006), Goumet et al. (2007), Ross et al. (2008), Hewett et al. (2009)
C21	Situation awareness modeling	Raggad (1997), Plessis and Toit (2006), Feng et al. (2009)
C22	Environmental awareness	Phillips-Wren et al. (2004), Koo et al. (2008), Gingsör Sen et al. (2008)
C23	Fuzzy decision-making	Metsotiotis et al. (2003), Zack (2007), Makropoulos et al. (2008), Wadhwa et al. (2009), Yu et al. (2009)

ysis allows you to study the properties of questionnaire and the items that make them up. The reliability analysis procedure calculates a number of commonly used measures of scale reliability and also provides information about the relationships between individual items in the measurement scale (Hair et al., 1998).

The fifth and sixth stages of research framework are based on "factor analysis" and are concentrated on extraction and identification of the BI Assessment criteria affecting the intelligence of AIS. Factor analysis is also known as a generic name given to a class of multivariate statistical methods whose primary purpose is to define the underlying structure in a data matrix. With factor analysis, the researcher can first identify the separate factors of the structure and then determine the extent to which each variable is explained by each factor. Once these factors and the explanation of each variable are determined, the two primary uses for factor analysis-summarization and data reduction-can be achieved. In summarizing the data, factor analysis derives underlying factors that, when interpreted and understood, describe the data in a much smaller number of concepts than the original individual variables (Hair et al., 1998). Evaluating the suitability of collected data, performing factor analysis and naming the extracted factors are different steps.

5 DISCUSSION

A questionnaire was designed and structured in three sections. Information related to the basic profile of the interviewees was requested at the beginning of the questionnaire.

In the second part, there were 23 questions designed to

measure their attitude, based on the BI Assessment Criteria listed in Table 2. The selected response was evaluated by a "Likert Scale" (Likert, 1974) and the responses could be: very strongly disagree, strongly disagree, disagree, no opinion, agree, strongly agree or very strongly agree. In other words, the second part of questionnaire measures their opinions the importance of each BI specification in terms of the intelligence Assessment Criteria of AIS.

The main targets of the sampling were accounting managers who are involved in systems efforts and decision-making.

5.1 Reliability analysis

With reliability analysis, you can get an overall index of the repeatability or internal consistency of the measurement scale as a whole, and you can identify problem items that should be excluded from the scale. The Cronbach's α is a model of internal consistency, based on the average inter-item correlation. The Cronbach's α (Likert, 1974) calculated from the 34 variables of this research was 0.941 (94 percent), which showed high reliability for designed measurement scale.

5.2 Data collection

The research targets were accounting managers who were involved in systems efforts and decision-making in organizations. The number of questionnaires sent out was 210 and the number returned was 176, which showed a return rate of 83 per cent.

5.3 Demographic profiles of interviewees

The demographic profile of interviewees who participate in the survey has been summarized in Table 2. The results show that most of the members (87.5 per cent) are male. Most of the interviewees (88.7 per cent) have a Bachelor of Science (BS) or a higher degree, as shown in Table 2.

TABLE 2
DEMOGRAPHIC PROFILES OF INTERVIEWEES

	Description	Number of interviewees	Percent
Gender	Male	154	87.5
	Female	22	12.5
	Sum	176	100
Educational degree	Under BS	20	11.4
	BS	83	47.2
	MS or higher education	73	41.5
	Sum	176	100
Seniority	Less than 5 years	7	4
	5 to less than 10 years	69	39.2
	10 to less than 15 years	64	36.4
	15 to less than 20 years	25	14.2
	20 years and above	11	6.2
	Sum	176	100

On the subject of decision-type, the majority of interviewees make semi-structured and unstructured decisions in their work. Table 2 also shows the seniority of the participants. As can be seen, 20.4 per cent have over 15 years of seniority, 36.4 per cent have 10-15 years, and 43.2 per cent have less than 10 years seniority.

5.4 Factor Extaction and Labeling

Factor analysis can be utilized to examine the underlying patterns or relationships for a large number of variables and to determine whether the information can be condensed or summarized in a smaller set of factors or components (Hair et al., 1998).

An important tool in interpreting factors is factor rotation. The term rotation means exactly what it implies. Specifically, the reference axes of the factors are turned about the origin until some other position has been reached. The un-rotated factor solutions extract factors in the order of their importance. The first factor tends to be a general factor with almost every variable loading significantly, and it accounts for the largest amount of variance. The second and subsequent factors are then based on the residual amount of variance. The ultimate effect of rotating the factor matrix is to redistribute the variance from earlier factors to later ones to achieve a simpler, theoretically more meaningful factor pattern. The simplest case of rotation is an orthogonal rotation, in which the axes are maintained at 90° (Hair et al., 1998).

In order to determine whether the partial correlation of the variables is small, the Kaiser-Meyer-Olkin was used to measure of sampling adequacy (Kaiser, 1958) and Bartlett's χ^2 test of Sphericity (Bartlett, 1950) before starting the factor analysis. The result was a KMO of 0.963 and Bartlett test p-value less than 0.05, which showed good correlation. The factor analysis method is "principle component analysis" in this research, which was developed by Hotteling (1935). The condition for selecting factors was based on the principle proposed by Kaiser (1958):

TABLE 3
ROTATED FACTOR ANALYSIS RESULTS

Factor	Initial Eigen values	Rotation Sums of Squared Loadings	
		Percentage of Variance	Cumulative percentage
1	16.114	21.366	21.366
2	3.173	18.366	39.732
3	1.850	14.836	54.568
4	1.457	8.755	63.324
5	1.348	6.025	69.349

Eigen value larger than one, and an absolute value of factor loading greater than 0.5. The 23 variables were grouped into five factors. The results can be seen in Table 3. Five factors had an Eigen value greater than one and the interpretation variable was 69.349 percent. The factors

were rotated according to Varimax rotation method.

To indicate the meaning of the factors, they have been given short labels indicating their content. "Analytical Decision-support", "Providing Integration with Environmental and Experimental Information", "Optimization Model", "Reasoning" and finally, "Enhanced Decision-making Tools" are the names which have been assigned to the extracted factors.

6 CONCLUSION

Enterprise systems like Accounting Information Systems (AIS) converts and store the data. Therefore, it is important to integrate decision-support into the environment of these systems. Business intelligence (BI) can be embedded in these enterprise systems to obtain competitive advantage.

This research confirmed the necessity to assess the Intelligence of Accounting Information Systems and demonstrated that this assessment can advance a decision-support environment. From a wide-ranging literature review, 23 criteria for BI assessment were gathered and embedded in the second part of the research. The interviewees selected the more important criteria from these 23 variables by assigning ranks to them. The research then applied factor analysis to extract the five factors for evaluation. These factors were "Analytical Decision-support", "Providing Integration with Environmental and Experimental Information", "Optimization Model", "Reasoning" and finally, "Enhanced Decision-making Tools". The authors believe that after this research, organizations can make better decisions for designing, selecting, evaluating and buying Accounting Information Systems, using criteria that help them to create a better decision-support environment in their work systems. Of course, further research is needed. One area is the design of expert systems (tools) to compare vendors' products. The other is the application these criteria and factors in a Multy Criteria Decision Making framework to select and rank AIS, financial and banking systems based on BI specification. The complex relationship between decision-making satisfaction of managers, and these factors should also be addressed in future research.

REFERENCES

- [1] Abernethy MA, Vagnoni E. Power, 2004. Organization design and managerial behaviour. *Account Organ Soc* ;29(3/4):207-25.
- [2] Alter, S., 2004. A work system view of DSS in its fourth decade. *Decision Support Systems*, 38(3), 319-327.
- [3] Anders Rom, Carsten Rohde, 2007. Management accounting and integrated information systems: A literature review, *International Journal of Accounting Information Systems* 8 40-68
- [4] Arnold V, Collier PA, Leech SA, Sutton SG. Impact of intelligent decision aids on expert and novice decision-makers' judgments. *Account Finance* 2004;44(1):1-26.
- [5] Azadivar, F., Truong, T. and Jiao, Y., 2009. A decision support system for fisheries management using operations research and systems science approach, *Expert Systems with Applications* 36: 2971-2978.
- [6] Bartlett, M.S., 1950. Test of significance in factor analysis, *British Journal of Psychology*, Vol. 3, 77-85.
- [7] Berson, A., & Smith, S., 1997. *Data warehousing, data mining, and OLAP*: McGraw-Hill, Inc. New York, NY, USA.
- [8] Birnberg, J. G., & Shields, J. F. (1989). Three decades of behavioral accounting research: A search for order. *Behavioral Research in Accounting*, 1, 23±74.
- [9] Bolloju, N., Khalifa, M. and Turban, E., 2002. Integrating knowledge management into enterprise environments for the next generation decision support, *Decision Support Systems* 33: 163-176.
- [10] Bose, R., 2009. Advanced analytics: opportunities and challenges, *Industrial Management & Data Systems*, Vol. 109, No. 2, 155-172.
- [11] Bouwens, J., & Abernethy, M. A. (2000). The Consequences of Customization on Management Accounting System Design. *Accounting, Organizations and Society*, 25(3), 221-241.
- [12] Courtney, J.F., 2001. Decision making and knowledge management in inquiring organizations: toward a new decision-making paradigm for DSS, *Decision Support Systems* 31: 17-38.
- [13] Damart, S., Dias, L. and Mousseau, V., 2007. Supporting groups in sorting decisions: Methodology and use of a multi-criteria aggregation/disaggregation DSS, *Decision Support Systems* 43: 1464-1475.
- [14] David JS, Gerard GJ,McCarthyWE. Design science: an REA-perspective on the future of AIS. In:ArnoldV, Sutton SG, editors. *Researching accounting as an information systems discipline*. Sarasota, FL, USA: American Accounting Association; 2002.
- [15] Delorme, X., Gandibleux, X. and Rodríguez, J., 2009. Stability evaluation of a railway timetable at station level, *European Journal of Operational Research* 195: 780-790.
- [16] Denna EL, Cherrington JO, Andros DP, Hollander AS. *Event-driven business solutions*. Homewood, IL, USA: Business One Irwin; 1993.
- [17] Eckerson, W., 2005. *Performance dashboards: Measuring, monitoring, and managing your business*. Wiley Press.
- [18] Elbashir, M., Collier, P. and Davern, M., 2008. Measuring the effects of business intelligence systems: The relationship between business process and organizational performance. *International Journal of Accounting Information Systems*, 9(3), 135-153.
- [19] Eom, S., 1999. Decision support systems research: current state and trends, *Industrial Management & Data Systems* 99/5 213±220.
- [20] Evers, M., 2008. An analysis of the requirements for DSS on integrated river basin management, *Management of Environmental Quality: An International Journal* Vol. 19 No. 1, pp. 37-53.
- [21] Fazlollahi, B. and Vahidov, R., 2001. Extending the effectiveness of simulation-based DSS through genetic algorithms, *Information & Management* 39: 53-65.
- [22] Feng, Y., Teng, T. and Tan, A., 2009. Modelling situation awareness for Context-aware Decision Support, *Expert Systems*

- with Applications 36: 455–463.
- [23] Galasso, F. and Thierry, C., 2008. Design of cooperative processes in a customer–supplier relationship: An approach based on simulation and decision theory, *Engineering Applications of Artificial Intelligence*.
- [24] Gao, S. and Xu, D., 2009. Conceptual modeling and development of an intelligent agent-assisted decision support system for anti-money laundering, *Expert Systems with Applications* 36: 1493–1504.
- [25] Geerts GL, McCarthy WE. An ontological analysis of the economic primitives of the extended REA enterprise information architecture. *Int J Account Inf Syst* 2002;3(1):1–16.
- [26] Gelinas UJ, Sutton SG, Hunton JE. *Accounting information systems*. 6th edition. Thomson, OH, USA: South-Western; 2005.
- [27] Ghoshal, S. and Kim, S.K., 1986. Building Effective Intelligence Systems for Competitive Advantage, *Sloan Management Review*, Vol. 28, No. 1, 49–58.
- [29] Gonnet, S., Henning, G. and Leone, H., 2007. A model for capturing and representing the engineering design process, *Expert Systems with Applications* 33: 881–902.
- [30] González, J.R., Pelta, D.A. and Masegosa, A.D., 2008. A framework for developing optimization-based decision support systems, *Expert Systems with Applications*.
- [31] Gottschalk, P., 2006. Expert systems at stage IV of the knowledge management technology stage model: The case of police investigations, *Expert Systems with Applications* 31: 617–628.
- [32] Goul, M. and Corral, K., 2007. Enterprise model management and next generation decision support, *Decision Support Systems* 43: 915–932.
- [33] Güngör Sen, C., Baraçlı, H., Sen, S. and Baslıgil, H., 2008. An integrated decision support system dealing with qualitative and quantitative objectives for enterprise software selection, *Expert Systems with Applications*.
- [34] Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W.C., 1998. *Multivariate Data Analysis*, Prentice-Hall, Upper Saddle River, NJ, 7-232.
- [35] Hedgebeth, D., 2007. Data-driven decision making for the enterprise: an overview of business intelligence applications, *The journal of information and knowledge management systems* Vol. 37 No. 4, 414-420.
- [36] Hemsley-Brown, J., 2005. Using research to support management decision making within the field of education, *Management Decision* Vol. 43 No. 5, pp. 691-705.
- [37] Hewett, C., Quinn, P., Heathwaite, A.L., Doyle, A., Burke, S., Whitehead, P. and Lerner, D., 2009. A multi-scale framework for strategic management of diffuse pollution, *Environmental Modelling & Software* 24: 74–85.
- [38] Hotteling, H., 1935. The most predictable criterion, *Journal of Educational Psychology*, Vol. 26, 139-142.
- [39] Kaiser, H., 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187-200.
- [40] Koo, L.Y., Adhitya, A., Srinivasan, R. and Karimi, I.A., 2008. Decision support for integrated refinery supply chains Part 2. Design and operation, *Computers and Chemical Engineering* 32: 2787–2800.
- [41] Koutsoukis, N., Dominguez-Ballesteros, B., Lucas, C.A. and Mitra, G., 2000. A prototype decision support system for strategic planning under uncertainty, *International Journal of Physical Distribution & Logistics Management*, Vol. 30 No. 7/8, 640-660.
- [42] Kwon, O., Kim, K. and Lee, K.C., 2007. MM-DSS: Integrating multimedia and decision-making knowledge in decision support systems, *Expert Systems with Applications* 32: 441–457.
- [43] Lamptey, G., Labi, S. and Li, Z., 2008. Decision support for optimal scheduling of highway pavement preventive maintenance within resurfacing cycle, *Decision Support Systems* 46: 376–387.
- [44] Lee, J. and Park, S., 2005. Intelligent profitable customers segmentation system based on business intelligence tools, *Expert Systems with Applications* 29: 145–152.
- [45] Li, D., Lin, Y. and Huang, Y., 2009. Constructing marketing decision support systems using data diffusion technology: A case study of gas station diversification, *Expert Systems with Applications* 36: 2525–2533.
- [46] Li, S., Shue, L. and Lee, S., 2008. Business intelligence approach to supporting strategy-making of ISP service management, *Expert Systems with Applications* 35: 739–754.
- [47] Likert, R., 1974. The method of constructing an attitude scale, in Maranell, G.M. (Ed.), *Scaling: A Sourcebook for Behavioral Scientists*, Aldine Publishing Company, Chicago, IL, 21-43.
- [48] Lin, Y., Tsai, K., Shiang, W., Kuo, T. and Tsai, C., 2009. Research on using ANP to establish a performance assessment model for business intelligence systems, *Expert Systems with Applications* 36: 4135–4146.
- [49] Loebbecke, C. and Huyskens, C., 2007. Development of a model-based netsourcing decision support system using a five-stage methodology, *European Journal of Operational Research*.
- [50] Lönnqvist, A. and Pirttimäki, V., 2006. The Measurement of Business Intelligence, *Information Systems Management*, 23:1, 32–40.
- [51] Makropoulos, C.K., Natsis, K., Liu, S., Mittas, K. and Butler, D., 2008. Decision support for sustainable option selection in integrated urban water management, *Environmental Modelling & Software* 23: 1448–1460.
- [52] Maria, F., 2005. Improving the utilization of external strategic information. Tampere University of Technology, Master of Science Thesis.
- [53] Marinoni, O., Higgins, A., Hajkowicz, S. and Collins, K., 2009. The multiple criteria analysis tool (MCAT): A new software tool to support environmental investment decision making, *Environmental Modelling & Software* 24: 153–164.
- [54] Metaxiotis, K., Psarras, J. and Samouilidis, E., 2003. Integrating fuzzy logic into decision support systems: current research and future prospects, *Information Management & Computer Security* 11/2, 53-59.
- [55] McCarthy WE. An entity-relationship view of accounting models. *Account Rev* 1979;54(4):667–86.
- [56] McCarthy WE. The REA accounting model: a generalized framework for accounting systems in a shared data environment. *Account Rev* 1982;57(3):554–78.
- [57] Nemati, H., Steiger, D., Iyer, L. and Herschel, R., 2002. Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data

- warehousing, *Decision Support Systems* 33: 143–161.
- [58] Nie, G., Zhang, L., Liu, Y., Zheng, X. and Shi, Y., 2008. Decision analysis of data mining project based on Bayesian risk, *Expert Systems with Applications*.
- [59] Noori, B. and Salimi, M.H., 2005. A decision-support system for business-to-business marketing, *Journal of Business & Industrial Marketing* 20/4/5: 226–236.
- [60] Nylund, A., 1999. Tracing the BI Family Tree. *Knowledge Management*.
- [61] O'zbayrak, M. and Bell, R., 2003. A knowledge-based decision support system for the management of parts and tools in FMS, *Decision Support Systems* 35: 487–515.
- [62] Petrini, M. and Pozzebon, M., 2008. What role is "Business Intelligence" playing in developing countries? A picture of Brazilian companies. In: Rahman, Hakikur (Eds.), *Data Mining Applications for Empowering Knowledge Societies*, IGI Global, 237–257.
- [63] Phillips-Wren, G., Hahn, E. and Forgionne, G., 2004. A multiple-criteria framework for evaluation of decision support systems, *Omega* 32: 323 – 332.
- [64] Phillips-Wren, G., Mora, M., Forgionne, G.A. and Gupta, J.N.D., 2007. An integrative evaluation framework for intelligent decision support systems, *European Journal of Operational Research*.
- [65] Pitty, S., Li, W., Adhitya, A., Srinivasan, R. and Karimi, I.A., 2008. Decision support for integrated refinery supply chains Part 1. Dynamic simulation, *Computers and Chemical Engineering* 32: 2767–2786.
- [66] Plessis, T. and Toit, A.S.A., 2006. Knowledge management and legal practice, *International Journal of Information Management* 26: 360–371.
- [67] Power, D. and Sharda, R., 2007. Model-driven decision support systems: Concepts and research directions, *Decision Support Systems* 43: 1044–1061.
- [68] Power, D.J., 2008. Understanding Data-Driven Decision Support Systems, *Information Systems Management*, 25:2, 149 — 154.
- [69] Quinn, N.W.T., 2009. Environmental decision support system development for seasonal wetland salt management in a river basin subjected to water quality regulation, *Agricultural Water Management* 96, 247 – 254.
- [70] Raggad, B.G., 1997. Decision support system: use IT or skip IT, *Industrial Management & Data Systems* 97/2: 43–50.
- [71] Ranjan, J., 2008. Business justification with business intelligence, *VINE: The journal of information and knowledge management systems*, Vol. 38, No. 4, 461-475.
- [72] Reich, Y. and Kapeliuk, A., 2005. A framework for organizing the space of decision problems with application to solving subjective, context-dependent problems, *Decision Support Systems* 41: 1 – 19.
- [73] Ross, J.J., Denar, M.A. and Mahfouf, M., 2008. A hybrid hierarchical decision support system for cardiac surgical intensive care patients. Part II. Clinical implementation and evaluation, *Artificial Intelligence in Medicine*.
- [74] Santhanam, R. and Guimaraes, T., 1995. Assessing the quality of institutional DSS, *European Journal of Information Systems* 4 (3).
- [75] Shang, J., Tadikamalla, P., Kirsch, L. and Brown, L., 2008. A decision support system for managing inventory at GlaxoSmithKline, *Decision Support Systems*.
- [76] Shi, Z., Huang, Y., He, Q., Xu, L., Liu, S., Qin, L., Jia, Z., Li, J., Huang, H. and Zhao, L., 2007. MSMiner—a developing platform for OLAP, *Decision Support Systems* 42: 2016–2028.
- [77] Shim, J., Warkentin, M., Courtney, J., Power, D., Sharda, R. and Carlsson, C., 2002. Past, present, and future of decision support technology, *Decision Support Systems* 33: 111–126.
- [78] Sutton SG, Arnold V. Foundations and frameworks for AIS research. In: Arnold V, Sutton SG, editors. *Researching accounting as an information systems discipline*. Sarasota, FL, USA: American Accounting Association; 2002.
- [79] Wadhwa, S., Madaan, J. and Chan, F.T.S., 2009. Flexible decision modeling of reverse logistics system: A value adding MCDM approach for alternative selection, *Robotics and Computer-Integrated Manufacturing* 25: 460–469.
- [80] Yu, L., Wang, S. and Lai, K., 2009. An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: The case of credit scoring, *European Journal of Operational Research* 195: 942–959.
- [81] Zack, M., 2007. The role of decision support systems in an indeterminate world, *Decision Support Systems* 43: 1664–1674.
- [82] Zhan, J., Loh, H.T. and Liu, Y., 2009. Gather customer concerns from online product reviews – A text summarization approach, *Expert Systems with Applications* 36: 2107–2115.
- [83] Zhang, X., Fu, Z., Cai, W., Tian, D. and Zhang, J., 2009. Applying evolutionary prototyping model in developing FIDSS: An intelligent decision support system for fish disease/health management, *Expert Systems with Applications* 36: 3901–3913.

Processing of Images Based on Segmentation Models for Extracting Textured Component

V M Viswanatha, Nagaraj B Patil, Sanjay Pande MB

Abstract— The method for segmentation of color regions in images with textures in adjacent regions being different can be arranged in two steps namely color quantization and segmentation spatially. First, colors in the image are quantized to few representative classes that can be used to differentiate regions in the image. The image pixels are then replaced by labels assigned to each class of colors. This will form a class-map of the image. A mathematical criteria of aggregation and mean value is calculated. Applying the criterion to selected sized windows in the class-map results in the highlighted boundaries. Here high and low values correspond to possible boundaries and interiors of color texture regions. A region growing method is then used to segment the image.

Key Words: Texture segmentation, clustering, spital segmentation, slicing, texture composition, boundry value image, median-cut.

1. INTRODUCTION

Segmentation is the low-level operation concerned with partitioning images by determining disjoint and homogeneous regions or, equivalently, by finding edges or boundaries. Regions of an image segmentation should be uniform and homogeneous with respect to some characteristics such as gray tone or texture. Region interiors should be simple and without many small holes. Adjacent regions of segmentation should have significantly different values with respect to the characteristic on which they are uniform. Boundaries of each segment should be simple, not ragged, and must be spatially accurate".

Thus, in a large number of applications in image processing and computer vision, segmentation plays a fundamental role as the first step before applying the higher-level operations such as recognition, semantic interpretation, and representation.

Earlier segmentation techniques were proposed mainly for gray-level images on which rather comprehensive survey can be found. The reason is that, although color information permits a more complete representation of images and a more reliable segmentation of them,

processing color images requires computation times considerably larger than those needed for gray-level images. With an increasing speed and decreasing costs of computation; relatively inexpensive color camera the limitations are ruled out. Accordingly, there has been a remarkable growth of algorithms for segmentation of color images. Most of times, these are kind of "dimensional extensions" of techniques devised for gray-level images; thus exploit the well-established background laid down in that field. In other cases, they are ad hoc techniques tailored on the particular nature of color information and on the physics of the interaction of light with colored materials. More recently, Yining Deng and B. S. Manjunath [1][2] uses the basic idea of separate the segmentation process into color quantization and spatial segmentation. The quantization is performed in the color space without considering the spatial distributions of the colors. S Belongie, et.al [3], in their paper present a new image representation, which provides a transformation from the raw pixel data to a small set of image regions that are coherent in color and texture space. A new method of color image

segmentation is proposed in [4] based on K-means algorithm. Both the hue and the intensity components are fully utilized.

2 COLOR QUANTIZATION

Color Quantization is a form of image compression that reduces the number of colors used in an image while maintaining, as much as possible, the appearance of the original. The optimal goal in the color quantization process is to produce an image that cannot be distinguished from the original. This level of quantization in fact may never be achieved. Thus, a color quantization algorithm attempts to approximate the optimal solution.

The process of color quantization is often broken into four phases .

- 1) Sample image to determine color distribution.
- 2) Select colormap based on the distribution
- 3) Compute quantization mapping from 24-bit colors to representative colors
- 4) Redraw the image, quantizing each pixel.

Choosing the colormap is the most challenging task. Once this is done, computing the mapping table from colors to pixel values is straightforward.

In general, algorithms for color quantization can be broken into two categories: Uniform and Non-Uniform. In Uniform quantization the color space is broken into equal sized regions where the number of regions N_R is less than or equal to Colors K . Uniform quantization, though computationally much faster, leaves much room for improvement. In Non-Uniform quantization the manner in which the color space is divided is dependent on the distribution of colors in the image. By adapting a colormap to the color gamut of the original image, it is

assured of using every color in the colormap, and thereby reproducing the original image more closely.

The most popular algorithm for color quantization, invented by Paul Heckbert in 1980, is the median cut algorithm. Many variations on this scheme are in use. Before this time, most color quantization was done using the popularity algorithm, which essentially constructs a histogram of equal-sized ranges and assigns colors to the ranges containing the most points. A more modern popular method is clustering using Octree,

3. Segmentation

The division of an image into meaningful structures, *image segmentation*, is often an essential step in image analysis. A great variety of segmentation methods have been proposed in the past decades. They can be categorized into

Threshold based segmentation: Histogram thresholding and slicing techniques are used to segment the image.

Edge based segmentation: Here, detected edges in an image are assumed to represent object boundaries, and used to identify these objects.

Region based segmentation: Here the process starts in the middle of an object and then grows outwards until it meets the object boundaries.

Clustering techniques: Clustering methods attempt to group together patterns that are similar in some sense.

Perfect image segmentation cannot usually be achieved because of *oversegmentation* or *undersegmentation*. In oversegmentation pixels belonging to the same object are classified as belonging to different segments.

In the latter case, pixels belonging to different objects are classified as belonging to the same object.

4. Design

Natural pictures are rich in color and texture. Texture segmentation algorithms require the estimation of texture model parameters which is difficult. The goal of this work is to segment images into homogeneous color-texture regions. The approach tests for the homogeneity of a given color-texture pattern, which is computationally more feasible than model parameter estimation. In order to identify homogeneity, the following assumptions are made:

- Each image contains homogeneous color-texture regions.
- The information in each image region can be represented by a set of few quantized colors.
- The colors between two neighboring regions are distinguishable

4.1 The Process of Segmentation

The segmentation is carried out in two stages. In the first stage, colors in the image are quantized to several representative classes to differentiate regions in the image. Then the image pixel values are replaced by their corresponding labels to form a image class-map that can be viewed as a special kind of texture composition. In the second stage, spatial segmentation is performed directly class-map. A few good existing quantization algorithms are used in this work.

The focus of this work is on spatial segmentation and can be summarized as:

- For image segmentation a parameter is calculated. This involves minimizing a cost associated with the partitioning of the image based on pixel labels.
- Segmentation is achieved using an algorithm. The notation of Boundary-images, correspond to measurements of

local homogeneities at different scales, which can indicate potential boundary locations.

- A spatial segmentation algorithm that grows regions from seed areas of the Boundary-images to achieve the final segmentation. Figure. 4.1 shows a schematic of the algorithm for color image segmentation.

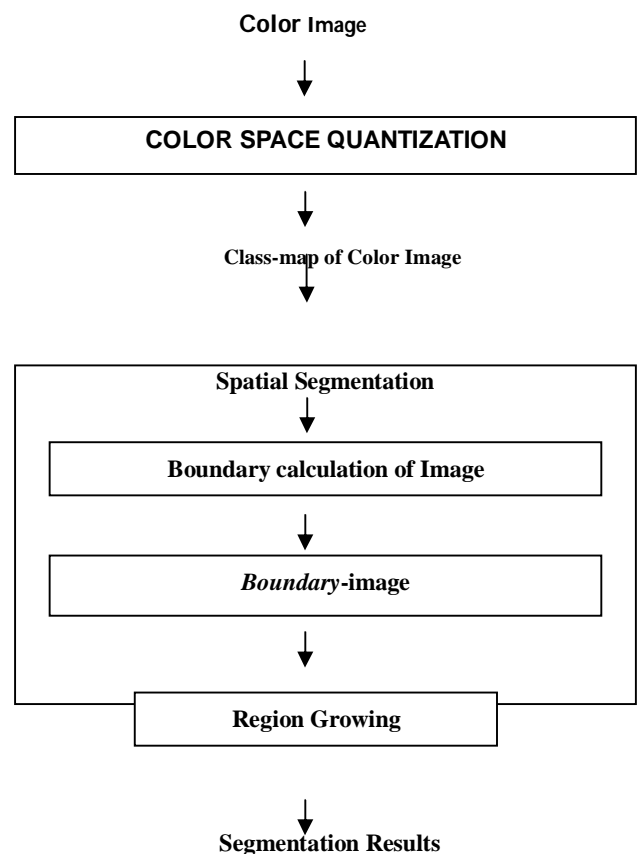


Fig.4.1: Schematic of Segmentation

4.2 The Parameter Calculation as a criterion for Segmentation

It is difficult to handle a 24-bit color images with thousands of colors. Image are coarsely quantized without significantly degrading the color quality. Then, the quantized colors are assigned labels. A color class is the set of image pixels quantized to the same color. The image pixel colors are replaced by their corresponding color class labels and this image is called a class-map. Usually,

each image region contains pixels from a small subset of the color classes and each class is distributed in a few image regions.

The class-map image is viewed as a special kind of texture composition. The value of each point in the class-map is the image pixel position, a 2-D vector (x, y) . Each point belongs to a color class.

Let Z be the set of all N data points in a class-map. Let $z = (x, y)$, $z \in Z$, and m be the mean,

$$m = 1/N \sum_{z \in Z} z$$

Suppose Z is classified into C classes, Z_i , $i = 1, \dots, C$. Let m_i be the mean of the N_i data points of class Z_i ,

$$m_i = 1/N_i \sum_{z \in Z_i} z$$

Let

$$S_T = \sum_{z \in Z} \|z - m\|^2$$

$$S_W = \sum S_i = \sum_{i=1}^C \sum_{z \in Z_i} \|z - m_i\|^2$$

S_W is the total variance of points belonging to the same class.

$$J = (S_T - S_W) / S_W$$

For the case of an image consisting of several homogeneous color regions, the color classes are more separated from each other and the value of J is large. An example of such a class-map is class-map 1 in Figure.4.2 and the corresponding J value is 1.7. Figure. 4.2 shows another example of a three class distribution for which J is 0.8, indicating a more homogeneous distribution than class-map 1. For class-map 2, a good segmentation would be

two regions. One region contains class '1' and the other one contains classes '2' and '0'.

Now J is recalculated over each segmented region instead of the entire class-map and define the average J_{avg} by

$$J_{avg} = 1/N \sum_k M_k J_k$$

where J_k represents J calculated over region k , M_k is the number of points in region k , N is the total number of points in the class-map, and the summation is over all the regions in the class-map. J_{avg} is the parameter used for segmentation. For a fixed number of regions, a better segmentation tends to have a lower value of J_{avg} . If the segmentation is good, each segmented region contains a few uniformly distributed color class labels and the resulting J value for that region is small. Therefore, the overall J_{avg} is also small. Notice that the minimum value J_{avg} can take is 0. For the case of class-map 1, the partitioning shown is optimal ($J_{avg}=0$).

Figure.4.2 shows the manual segmentation results of class-maps.

Since the global minimization of J_{avg} for the entire image is not practical, J , if applied to a local area of the class-map, indicates whether that area is in the region interiors or near region boundaries

$$J=1.72$$

$$J=0.855$$

$$J1 = 0, J2 = 0, J0 = 0$$

$$J1 = 0, J\{2,0\} = 0.011$$

$$J_{avg} = 0$$

$$J_{avg} = 0.05$$

11111	0000
11111	0000
11111	0000
11111	0000
11111	0000
11111	0000
1111	22222
1111	22222
1111	22222
1111	22222

11111	2020
11111	0202
11111	2020
11111	0202
11111	2020
1111	20202
1111	02020
1111	20202
1111	02020

Figure 4.2: Segmented class-maps and their corresponding J values.

Now an image is constructed whose pixel values correspond to these J values calculated over small windows centered at the pixels. These are referred as *Boundary-values Images* and the corresponding pixel values as local J values. The higher the local J value is, the more likely that the corresponding pixel is near a region boundary. The *Boundary-image* contains intensities that actually represent the region interiors and region boundaries, respectively. Windows of small size are useful in localizing the intensity/color edges, while large windows are useful for detecting texture boundaries. Often, multiple scales are needed to segment an image. In this implementation, the basic window at the smallest scale is a 9×9 window. The smallest scale is denoted as scale 1. The window size is doubled each time to obtain the next larger scale.

The characteristics of the *Boundary-images* allow us to use a region-growing method to segment the image. Initially the entire image is considered as one region. The segmentation of the image starts at a coarse initial scale. It then repeats the same process on the newly segmented regions at the next finer scale. Region growing consists of determining the seed points and growing from those seed locations. Region growing is followed by a region merging operation to give the final segmented image.

5. Implementation

The implementation of segmentation is carried out using JDK 1.5. and JAI 1.3 (Java Advanced Imaging 1.3) on windows platform.

The module ColorQuantizer implements a GUI that takes the number of colors the image should be quan-

tized into along with the algorithm to choose from between Median-Cut, NeuQuant, Oct-Tree.

The module Segmentor gets arguments regarding the initial window size and maximum number of iterations. Here the segmentation parameter is calculated and boundary image is created. It scans the color vector in boundary image and calculates initial segmented image. Then for the number of iterations mentioned it segments images repeatedly by converging the values. The accuracy of segmentation depends on the number of iterations the image is segmented.

The module for Region Growing implements a simple region growing algorithm. It runs a classic stack-based region growing algorithm. It finds a pixel that is not labeled, labels it and stores its coordinates on a stack. While there are pixels on the stack, it gets a pixel from the stack (the pixel being considered), Checks its neighboring pixels to see if they are unlabeled and close to the considered pixel, if are, label them and store them on the stack. Repeats this process until there are no more pixels on the image.

Figure 5.1: Snapshot of Selected Image and Segmented Image (NeuQuant,12)

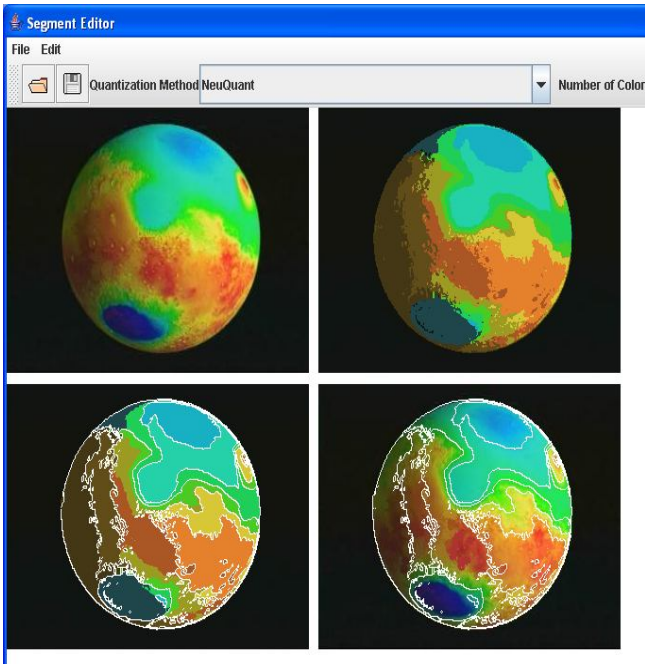
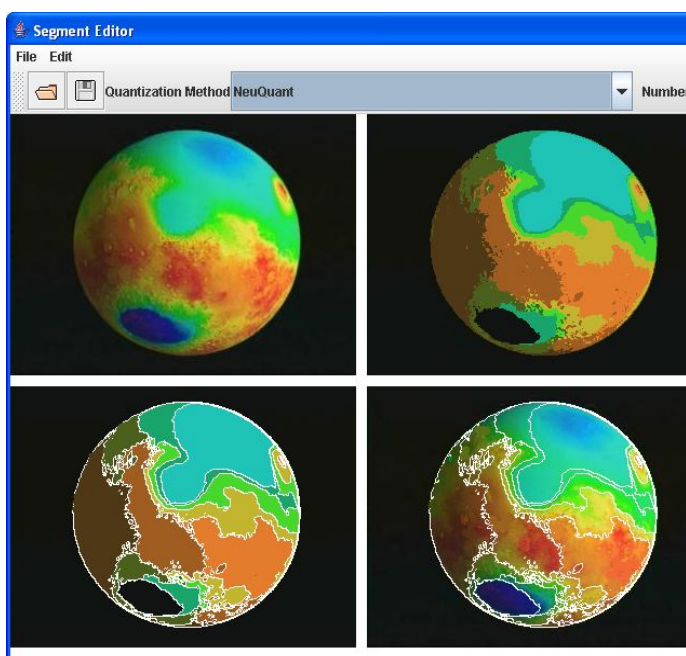


Figure 5.2: Snapshot of Selected Image and Segmented Image (NeuQuant,16)



In Figures 5.1 thru 5.3 ,we can see the border displayed on the original image, border displayed on the segmented image, original image and segmented image , shown in top left to bottom right in that order

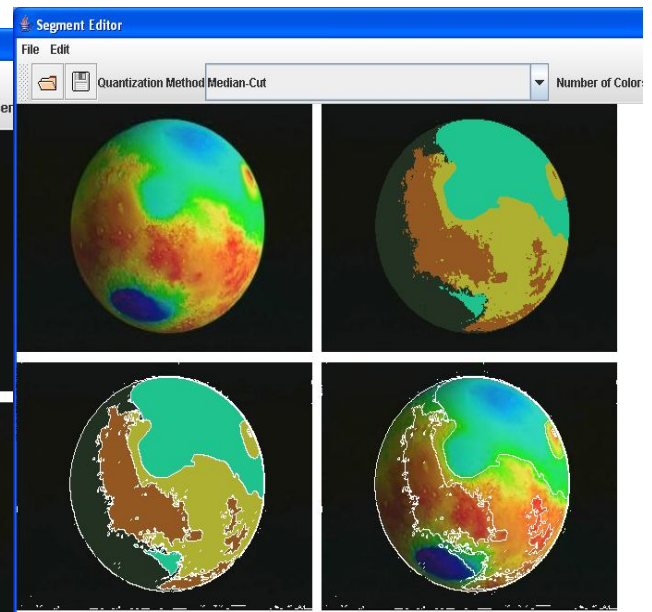
It is observed that changing of the parameters

also changes the result. The number of colors for quantization chosen and the number of iterations changes the number of regions in the segmented image.

6. Conclusion

The color image is quantized and then is segmented with a focus on spatial segmentation. The threshold values are calculated that is used for the segmentation at various scales. At each step the segments are identified through the boundaries and valleys that are identified with the threshold values that are low near the boundary. The accuracy of segmentation depends on the algorithm used for quantization and the number of iterations.

Figure 5.3: Snapshot of Selected Image and Segmented Image (Median-Cut,8)



the segmentation is over the image is displayed with appropriate borders to differentiate between the segments. The images having different colors in the adjacent regions are the best candidates for the segmentation using the proposed method.

However, it does not handle pictures where smooth transition takes place between adjacent regions and there is no clear visual boundary. For instance, the color of a sunset sky can vary from red to orange to dark in a very smooth transition.

7. Future Work

The algorithm can be modified to find out the number of colors adaptively for quantization instead of taking it as a parameter.

The algorithm can be extended to handle images with smooth transitions between adjacent regions. The method can be extended to segment the color-texture regions in video data.

ACKNOWLEDGEMENTS

The authors like to thank Singhania University, Rajasthan and VVIT Institute of Technology, Mysore.

BIBLIOGRAPHY

- [1] Y. Deng, B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Aug 2001.
- [2] Y. Ma and B. S. Manjunath, "EdgeFlow: A technique for boundary detection and segmentation", *IEEE Transactions on* Vol. 9, pp. 1375-88, August 2000.
- [3] S. Belongie, C. Carson, H. Greenspan and J. Malik, "Color and texture based image segmentation using EM and its application to content based image retrieval", *International Conference on Computer Vision*, pp. 675-682, 1998.
- [4] Pappas, T.N., "An adaptive clustering algorithm for image segmentation", *International Conference on Pattern recognition*, Vol 3, pp 3619, 2000.
- [5] H.D. Cheng, X.H. Jiang, Y. Sun and Jingli Wang, "Color Image Segmentation: Advances and Prospects", *IEEE Transactions on Pattern Recognition*, Vol 34, pp 2259- 2281, December 2001.
- [6] D.K. Panjwani and G. Healey, "Markov random field models for unsupervised segmentation of textured color images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 17, no.10, p. 939-54, 1995.
- [7] Heng-Da Cheng Ying Sun, "A hierarchical approach to color image segmentation using homogeneity", Vol 9, pp 2071- 2082.
- [8] Jianbo Shi, Jitendra Malik, "Normalized cuts and image segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 22, pp 888-905, August 2000.
- [9] O'Reilly - Safari Books Online - Java 2D Graphics.
- [10] <http://java.sun.com/products/java-media/jai/forDevelopers/>
- [11] *Digital Image Processing*, Rafael C. Gonzalez, Richard E. Woods, Pearson Education, Addison Wesley Longman
- [12] *Digital Image Processing, a practical introduction using Java*, Nick Efford, Pearson Education

An Adaptive and Efficient XML Parser Tool for Domain Specific Languages

W. Jai Singh, S. Nithya Bala

Abstract— XML (eXtensible Markup Language) is a standard and universal language for representing information. XML has become integral to many critical enterprise technologies with its ability to enable data interoperability between applications on different platforms. Every application that processes information from XML documents needs an XML Parser which reads an XML document and provides interface for user to access its content and structure. However the processing of xml documents has a reputation of poor performance and a number of optimizations have been developed to address this performance problem from different perspectives, none of which have been entirely satisfactory. Hence, in this paper we developed a Fast Parser tool for domain specific languages. In this we can execute Parser user friendly without having any constraints.

Index Terms—XML, Parser Tool, Document Object Model, SAX, XML Document, Document Validation.

1 INTRODUCTION

The XML (eXtensible Markup Language) is now widely adopted within (networked) applications. Due to its flexibility and efficiency in transmission of data, XML has become the emerging standard of data transfer and data exchange across the application and Internet [1]. XML has potential as a back end solution as well as a marvelous standard for re-designing databases and other content. XML has become integral to many critical enterprise technologies with its ability to enable data interoperability between on different platforms. With various conversion tools now emerging in the marketplace, XML can be used to bridge between different applications. It's standard-based, set forth a design for structuring future content [1], [2].

XML delivers key advantages in interoperability due to its flexibility, expressiveness and platform-neutrality. As XML has become a performance-critical aspect of the next generation of business computing infrastructure. Tomorrow's computers will have more cores rather than exponentially faster clock speeds, and software will increasingly have to rely on parallelism to take advantage of this trend.

Every application that processes information from XML documents needs an XML Parser which reads an XML document and provides interface for user to access its content and structure. An XML parser facilitates in simplifying the process of manipulating XML documents. There are mainly two challenges for generic XML parsers. One is that the code size of the XML Parsers is restricted because of the limitation of memory. The other is the run time adaptability of XML Parsers is required due to the diversity of applications in terms of their dependency on XML syntax set.

Several efforts have been made to address the parsing and validation performance through the use of grammar based parser generation by leveraging XML schema language such as DTD (Document Type Definition), XML schema at compile time. DOM (Document Object Model) and SAX (Simple API for XML) are the two most widely used XML parsing models, none of which has been entirely satisfactory [2], [5].

A parser can read the XML document components via Application Programming Interfaces (APIs) in two approaches. For stream-based approach such as SAX (also known as event-based parser) and tree-based approach such as DOM. DOM (Document Object Model) and SAX (Simple API for XML) are the two most widely used XML parsing models, none of which has been entirely satisfactory [2], [3], [5], [7]. A brief description of them is given as follows.

DOM is a tree-based interface that models an XML document as a tree of various nodes. The main advantage of this parse method is that it supports random access to the document. DOM parsers create a node object for each node that precisely models all the structure and content information [2], [3], [5]. DOM is an easy way to work with XML. However, DOM parsers take too much time and memory, making them unavailable for large XML documents. Moreover, no actual work can be done before completely parsing XML, which introduces significant delay and is unacceptable in enterprise applications.

SAX is an event-based parsing model that reads an XML document from beginning to end. Each time it encounters a syntax construction; it generates an event and notifies the application [2], [5], [7]. It does not preserve the structure and content information in memory, thus saving a large amount of memory space. Unfortunately, they lack the ability of random access and are forward only, which limits their use to a very small scope. SAX is memory efficient but writing a SAX parser is complex.

Several methods were presented to improve XML parsing from different viewpoints. There are a number of approaches trying to address the performance bottleneck of XML parsing. The typical software solutions include the pull-based

- W. Jai Singh, Assistant Professor, Department of MCA, Park College of Engineering and Technology, Coimbatore, Tamil Nadu, India – 641 659.
- S. Nithya Bala, PG Scholar, Department of MCA, Park College of Engineering and Technology, Coimbatore, Tamil Nadu, India – 641 659

parsing [9], lazy parsing [10] and schema-specific parsing [4].

Su Cheng Haw and G. S. V. Radha Krishna Rao have presented a model called "Comparative Study and Benchmarking on XML Parser". In that, they compare the xerces and .NET parsers based on the performance, memory, usage and so on [2]. Giuseppe Psaila have been developed a system called "Loosely coupling Java algorithms and XML Parsers". In that, he conducted a study about the problem of coupling java algorithms with XML parsers. Su-cheng Haw and Chien-Sing Lee have been presented a model called "Fast Native XML Storage and Query Retrieval". In that, they proposed the INLAB2 architecture comprises of five main components namely XML parser, XML encoder, XML indexer, Data manager and Query processor [10]. Fadi El-Hassan and Dan Ionescu presented "An efficient Hardware based XML parsing techniques". In that, they proposed hardware based solutions can be an obvious choice to parse XML in a very efficient manner.

The existing XML Parsers spend a large amount of time in tokenizing the input. To overcome all the drawbacks, here we have developed a new Fast Parser tool for domain specific languages. Though careful analysis of the operations required for parsing and validation, we are using hash table to store element information, this will enhance the speed of accessibility while searching for an element. More over we are using regular expressions to search for the tags and attributes, this will enhance the speed while reading XML contents.

2 Fast XML Parser

To parse an XML document in software, the processing sequence starts by loading the XML document, then reading its characters in sequence, extracting elements and attributes and then validating the XML document, writing parsed information and finally reading the resulting parsed data. Our initial approach separates the process of reading the XML document and stores the contents in to the hash table using regular expressions. Fig.1 shows the architecture of the fast XML parser tool.

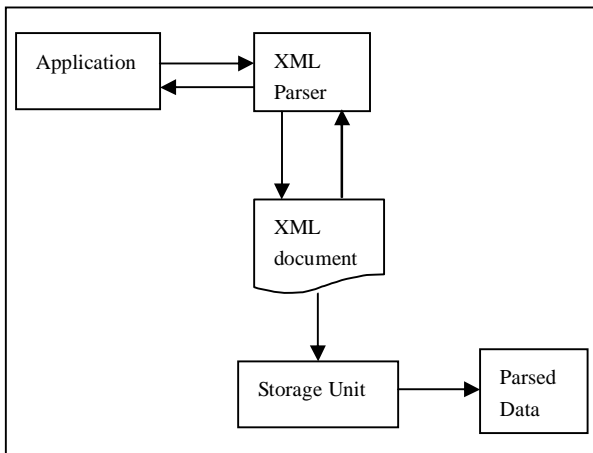


Fig.1: XML Parser Architecture

The Fast XML Parser Tool contains four modules. First, load an XML document in an application. Second, reading an XML document. Third, Writing an XML document into the application. Finally, knowledge based search of an XML docu-

ment. The Fast XML Parser Tool is as follows:

Load an XML document: Before an XML document can be accessed and manipulated, it must be loaded into an XML Parser. The XML Parser reads XML document and converts it into a meaningful format. The job of the XML Parser is to make sure that the document meets the defined structure and constraints. The validation rule for any particular sequence of character data is determined by the type of definition for its enclosing element. Fig.2 shows the sample XML document.

Algorithm

- Step 1: Read XML File
- Step 2: Search for start tag using regular expression '<...>
- Step 3: If a start tag found then
- Step 4: Add attributes to hash table
- Step 5: Search or end tag using regular expression '</...>
- Step 6: Add element to hash table
- Step 7: End if
- Step 8: Repeat step 3 to 7 until End Of File
- Step 9: Verify and validate each element in hash table
- Step 10: End XML Parser

Reading an XML document: Reading an existing load XML file. It provides the createXMLReader function that returns an implementation of the XMLReader interface. It reads the root element first and reads the sub element and corresponding data. Finally creates the XML document and sends it to the user application as a XML document.

Writing an XML document: Read the XML document and separates the content and writes it to the corresponding hash table such as root element, sub element, attributes and data. It provides the CreateXMLWriter function to return an implementation of the XMLWriter function.

Knowledge based search: Search a particular element or content from an XML document by our fast XML Parser tool. Initially it checks the content in storage unit using hash key that is root element. If it is available then it goes to the sub element and corresponding data and displays it as output. If it is not available means it terminates the search.

```

    <?xml version="1.0" encoding="UTF-8"?>
    <!-- Product data is the root element-->
    <PRODUCTDATA>
    <PRODUCT PROID="P001">
    <PRODUCTNAME>Barbie Doll</PRODUCTNAME>
    <DESCRIPTION>This is a toy for children in the age group of 5-10 years</DESCRIPTION>
    <PRICE>$120</PRICE>
    <QUANTITY>12</QUANTITY>
    </PRODUCT>
    </PRODUCTDATA>
    
```

Fig.2: Sample XML Document

3 EXPERIMENTAL RESULTS AND DISCUSSION

According to the experiment and considering the balance between memory and computing consumption, our parser tool is appropriate for parsing XML document with less memory. When XML documents size increases and gets larger, the time difference will increase accordingly.

We made some assumptions to facilitate our estimation. One of the most noticeable assumptions is that we simply ignore the intensive garbage collection operations in DOM. Garbage collection is actually very time-consuming, our parser tool can avoid it, because we does not recycle the unused memory space when parsing finishes. If we take this into account, the time difference would be even larger. So, we can draw the conclusion that our parser gains great advantage over DOM especially for large and complicated XML documents.

3.1. Testing Results

The testing result shows that great improvement has been achieved over present XML parsers, both in terms of memory computation and parsing performance. The tests are performed on our XML parser tool and MSXML. MSXML is a typical DOM parser which is widely used in enterprise applications. Similar tests are performed on other open-source XML parsers too, like Xerces-C++, Jaxen and Xalan. For clarity, we only present the testing results between our XML parser tool and MSXML. The testing codes are run in a HP G61 with 2.80GHz Intel(R) Core2 Due CPU and 4GB of RAM.

3.2. Memory Consumption

In this part, we test the ratio of memory space needed to parse an XML document to the original XML document size. The memory usage when many 1KB and 5KB XML documents are stored and accessed. The result shows in fig.3 that the memory consumption is acceptable in such an environment.

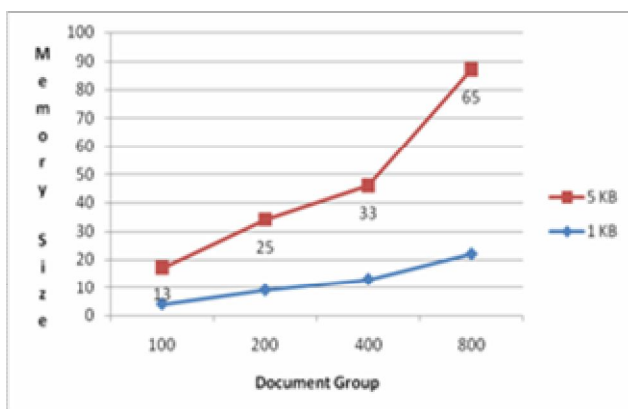


Fig.3: Memory Consumption Results

3.3 Parsing Performance

The results of parsing throughput per second are shown in figure. Note that our XML parser tool gets the lowest throughput, thus leading to low parsing performance. Figure 4 shows the cumulative parsing times for a group of similar XML documents. The X-axis is the file size of the XML documents that were parsed.

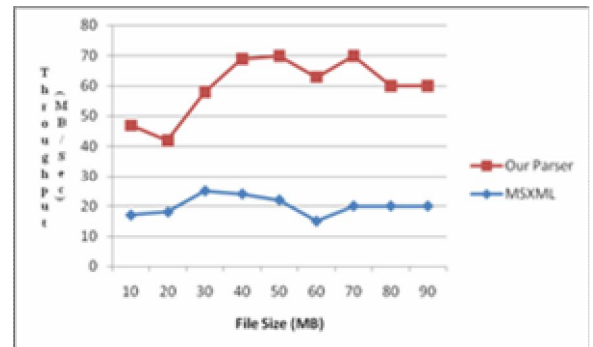


Fig.4: Parsing Throughput Results

3.4 Results Analysis

From above testing results we can see that our XML parser tool consistently outperforms MSXML in both memory consumption and parsing performance. Our XML parser tool consumes memory about approximately the same size of original XML document, while MSXML takes 3-4 times. Our XML parser tool deliveries up to 20-35MB/sec sustained throughput, while the average parsing throughput of MSXML is only 12MB/sec. Note that for MSXML, parsing throughput declines sharply when XML documents become large, which verifies the conclusion that too much creation and destroy of node objects in DOM can slow down parsing performance considerably.

4 CONCLUSION

From a proper analysis of positive points and constraints on the XML Parser tool, it can be safely concluded that the product is a highly efficient. This XML Parser tool is working properly and meeting to all user requirements. This tool can be easily plugged in domain specific languages. Our Parser tool can improved the XML parsing performance significantly and scales well. The computing efficiency will be improved in the future work.

5 REFERENCE

- [1] "eXtensible Markup Language", <http://www.w3.org/xml>. 10th Feb 2011.
- [2] Su Cheng Haw and G. S. V. Radha Krishna Rao, "A Comparative Study and Benchmarking on XML Parsers", *International Conference on Computer science and Technology*, IEEE Computer Society,

pp.321-325, Feb-2007.

- [3] Shiren Ye and Tat-Seng Chua, "Learning Object Models from Semistructured Web Documents", *IEEE Transactions on Knowledge and Data Engineering*, pp. 334-339, 2006.
- [4] Zhenghong Gao, Yinfei Pan, Ying Zhang and Kenneth Chiu, "A High Performance Schema-Specific XML Parser", *Third IEEE International Conference on e-Science and Grid Computing*, IEEE Computer Society, pp.245-252, 2007.
- [5] A. Waheed, J. Ding, "Benchmarking XML Based Application Oriented Network infrastructure and Services", *International Symposium on Applications and the Internet*, IEEE Computer Society, 2007.
- [6] Wei Zhang and Robert A. van Engelen, "An Adaptive XML Parser for Developing High-Performance Web Services", *Fourth IEEE International Conference on eScience*, IEEE Computer Society, 2008.
- [7] Yunsong Zhang Lei Zhao* Jiwen Yang Liying Yu, "NEM-XML: A Fast Non-extractive XML Parsing Algorithm", *Third International Conference on Multimedia and Ubiquitous Engineering*, IEEE Computer Society, 2009.
- [8] Zhou Yanming and Qu Mingbin, "A Run-time Adaptive and Code-size Efficient XML Parser", *30th Annual International Computer Software and Applications Conference*, IEEE Computer Society, 2006.
- [9] Giuseppe Psaila, "Loosely Coupling Java Algorithms and XML Parsers: a Performance-Oriented Study", *22nd International Conference on Data Engineering Workshops*, IEEE Computer Society, 2006.
- [10] Su-Cheng Haw and Chien-Sing Lee, "INLAB2: Fast Native XML Storage and Query Retrieval", *3rd International Conference on Intelligent System and Knowledge Engineering*, IEEE Computer Society, pp.44-49, 2008.
- [11] Fadi El-Hassan and Dan Ionescu, "SCBXP: An efficient hardware-based XML parsing technique", IEEE Computer Society, pp.45-50, 2009.
- [12] Wei Lu , Kenneth Chiu and Yinfei Pan, "A Parallel Approach to XML Parsing", *Grid Computing Conference* , pp.223-230, 2006.
- [13] Gang WANG, Cheng XU, Ying LI, Ying CHEN, "Analyzing XML Parser Memory Characteristics: Experiments towards Improving Web Services Performance", *IEEE International Conference on Web Services*, IEEE Computer Society, 2006.

ISSN 2229-5518



9 772229 551823



04