DESIGN OF ULTRA LOW POWER SYSTEMS BASED ON ADAPTIVE VOLTAGE OVER SCALING

Mr. M. Zaheer Ahmed[1], Mr. S. Lakshmi Kanth Reddy[2] ,Mr. A. Rajendra babu[3] and Dr. S. Nagaraju Rao[4]

[1]Department of ECE, Brindavan Institute of Tech. & Science.,JNTUA, zaheer.mtech@gamil.com). Kurnool, A.p.

[2]Department of ECE, Global College of Engineering &Technology.,JNTUA, kanth.srec@gamil.com). Kadapa, A.p.

[3]Department of ECE, B.C.E.W.,JNTUA, arajendrababuavula @gamil.com). Kadapa, A.p.

[4]Department of ECE G Pulla Reddy College of Engineering &Technology.,JNTUA, Kurnool, A.p.

ABSTRACT

This paper proposes a design methodology for voltage overscaling (VOS) of ultra-low-power systems. This paper first proposes a probabilistic model of the timing error rate for basic arithmetic units and validates it using both simulations and silicon measurements of multipliers in 65-nm CMOS.The model is then applied to a modified K-best decoder that employs error tolerance to reveal the potential of the framework. With simple modifications and timing error detection-only circuitry, the conventional K-best decoder improves its error tolerance in child node expansion modules by up to 30% with less than 0.4-dB SNR degradation.With this error tolerance, the supply voltage can be over scaled by 12.1%, leading to 22.5% energy savings.

Key Words: *Error-tolerant system, K-best decoder, low power circuit design, voltage overscaling.*

1. INTRODUCTION

Along With Process Improvements, Voltage Scaling Has Also Been Applied In A Wide Range Of Applications To Further Reduce Power Consumption. However, It Also Increases Stage Delay, Translating To Significant Leakage Energy Per Cycle In The Near And Subthreshold Regimes. Therefore, A Lower Bound On Energy Per Operation Is Reached In These Operating Regimes [1], [2]. However, The Worst-Case Critical Path In A Design Is Not Always Exercised, And Supply Voltage Can Be Overscaled While Maintaining A Fixed Performance If The System Can Tolerate Timing Errors. Voltage Overscaling (Vos) Enables Improved Energy Efficiency Beyond The Aforementioned Lower Bound If The Timing Correctness Assumption Is Relaxed. To circuit techniques [3], [4] have been proposed to detect and correct timing errors. However, the energy overhead of error correction eventually exceeds the energy savings from voltage scaling as the timing error rate rises. On the other hand, some digital signal processing (DSP) systems having in algorithmic error tolerance or can be modified to achieve error tolerance without significant quality degradation, as shown in [5] and [6]. these types of systems benefit more from vos compared to error-free error-correction to design a oltage-over scaled error tolerant system,we need to understand the tradeoff between energy savings and quality degradation during the design phase and select theoptimal design point. therefore, an accuratetimingerror model is a key enabler to vos-based design. oneapproachto building such a model is exhaustive search with simulation tools. however, the required design effort increases as system complexity grows. a preferred alternative would be a impleprobabilistic timing error model that provides reasonable accuracy, which will make the design of large dsp systems targeted at extreme energy efficiency feasible.thispaper proposes a practical design framework using an analytical timing error distribution model. starting from the analysis of a simple ripple carry adder (rca), we derive a normallydistributed error model for more complex circuits.weshow that the model fits reasonably well with typical circuit building blocks using simulation and measurement results. we then apply the proposed model to an error-resilient k-best decoder.the implemented design shows 22.5% energy savings beyond error-free computation at the expense of a small snr degradation of less than 0.4 db.

2. DESIGN APPROACHES FOR VOLTAGE-OVER SCALED SYSTEMS

The computation quality of many DSP systems can be measured by SNR. Generally, bly to nce

$$E_T = \sum_{N=T/d_{fa}}^{32} p_N \approx 0.5^{T/d_{fa}}$$

target. In [7], the authors perform extensive simulations and in-vestigate the benefit of VOS in a finite impulse response (FIR) filter, focusing on the error magnitude metric without architectural consideration. Therefore, the SNR performance degrades sharply as the supply voltage is scaled, limiting the energy improvements. More advanced approaches that incorporate error-correction schemes were also proposed. In [6] and [8], a noise-reduction unit suppresses noise generated from a voltage-overscaled FIR filter.In [10], a residue number system is applied along with other techniques to suppress errors in the voltage overscaled domain. However, overhead from the additional noise-reduction unit limits the benefit of VOS. Design approaches to enhance or estimate the effect of VOS have been previously proposed. Reference [11] suggests a design flow to redistribute slack for maximizing the amount of VOS. In [9], input and error statistics are analyzed at the PVT corners. These methods require extensive the probability ET that a timing error occurs for 32-bit adder is given by determine the degree of improvement. Reference [12] proposes an enhanced tool to obtain dynamic behavior of circuits. In [13],clock-skew scheduling is performed based on the importance of each signal to improve voltage scalability. Although this provides a simple design approach for voltage overscaled systems, process control of clock skew in ultra-lowpower designs is challenging given that the reduced operating voltages lead to heightened PVT simpler prediction models for timing errors using mathematical pproaches. However, the model in [13] has not been applied to arithmetic units beyond RCAs. Also, the approach in targets general purpose processors and therefore cannot consider the intrinsic properties of the DSP modules, as is done in this The exercise with simple RCA helps uncover the nature of timing errors for a more intuitive model. The worst-case critical path of a RCA is well known to be the carry propagation path from the where p is the probability that the carry input of the full adder propagates to the next stage. Then, the error rate is given by the probability that the length of the activated critical path is larger than the clock period. For p is 0.5 and clock period T, the probability ET that a timing error occurs for 32-bit adder is given by work. Finally, no proposed voltage-overscaled designs have been demonstrated in silicon.

3. ERROR ANALYSIS IN VOLTAGE-OVERSCALED SYSTEMS

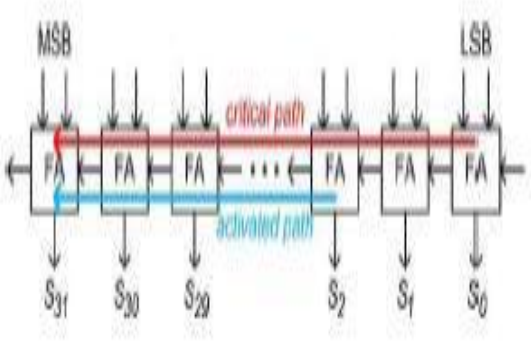
With conventional techniques discussed above, upon design modification, e.g., by changing the pipeline stage assignment be performed again with mostly different variables, making it challenging to find the optimal design point. Instead, we propose a simple timing error model that can be applied to general complex circuits with acceptable accuracy to provide design guidance. In this section,we propose a gaussian distribution-based timing error model by observations from RCA and generalize it. Finally, it is verified with fabricated Baugh-Wooley multipliers.

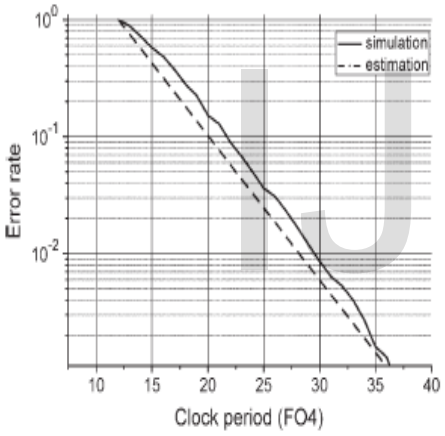
A: RCA Starting Point

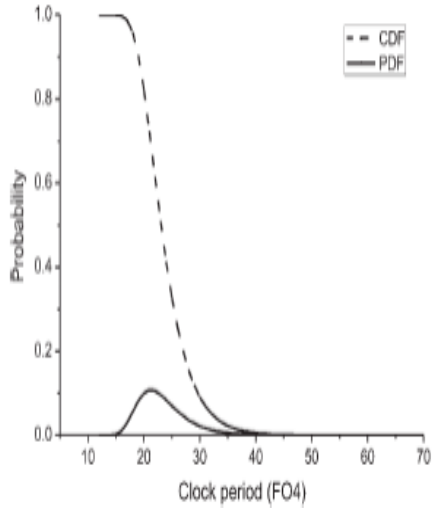
The adder is a basic element in typical DSP systems, and even larger modules such as multipliers are frequently implemented with multiple adders. By first finding an accurate error model for adders, we can then seek to extend it to the analysis of larger and more complex modules. The exercise with simple RCA helps uncover the nature of timing errors for a more intuitive model. The worst-case critical path of a RCA is well known to be the carry propagation path from the LSB to MSB, as shown in Fig. 1. However, the full critical path is activated infrequently, and this observation allows us to calculate the probability of timing errors across the input vector space. Since the path from LSB to MSB is the longest, we assume that the packet error rate can be approximated by the MSB error rate. Then, the probability that the length of the carry propagation path culminating at the MSB is exactly N in terms of the number of full adders is given by

$$p_N = p^N(1-p)$$

critical paths following independent delay distributions, the error rate at the clock period T is given by







v . 2
s on
a ng
t ror
r ore
c nd

tne error pattern starts to benave more probabilistically in contrast to the relatively deterministic model in (2). Given a system with N

where Ek,T is the error rate of the kth critical path. To simplify, we assume that all critical paths have distributions identical to (2) in order to represent a complex system with multiple critical paths of the same delay. Then, we obtain the

$$E_{system;T} = 1 - \prod_{k=1}^{N}(1 - E_{k,T})$$

slack shrinks, error rate behavior is similar to a Gaussian distribution. Although the PDF in Fig. 3 is slightly skewed rightward, we can still approximately fit it using a Gaussian distribution because this is pessimistic in modeling voltage-overscaled systems since the long tail exhibited in Fig. 3 will push the zero-error point farther right and lead to more achievable gains through VOS (i.e., larger margins to be exploited via VOS) than would be found with a traditional Gaussian decay in the delay PDF.."

B. Generalized Critical Path Delay Model

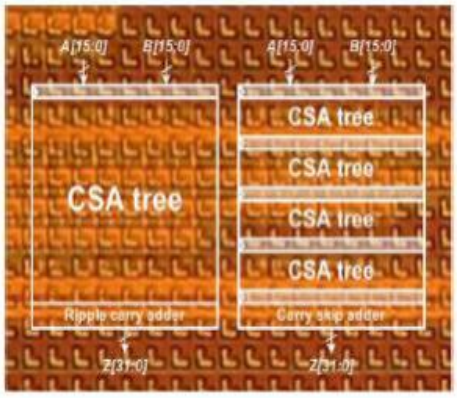
As the supply voltage reduces at a fixed clock speed, or clock period reduces at a fixed voltage, tim slack of the module also shrinks until a p d the cl iming di will impact the error rate. However, here we focus on the primary critical paths and assume they dominate timing errors at relatively low error rates for simplicity and practicality. If critical path delay is modeled as a Gaussian random variable over the input vector space with worst-case delay of T FO4, it can be viewed as an inverter chain of length T. Each inverter delay also follows the Gaussiandistribution with mean μ and standard deviation σ. Therefore, the delay of the entire critical path becomes the sum of random variables Where Dpath andDinv,k are the delays of the entire critical path and single inverter, respectively. Then, the timing error rate is the

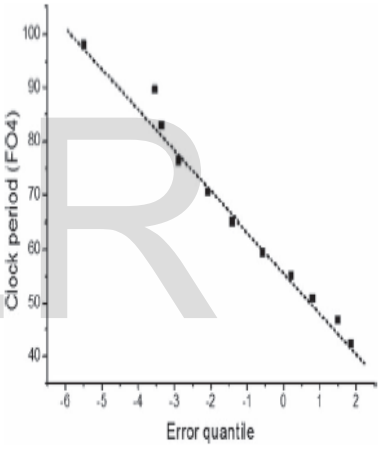
$$D_{path} = \sum_{k=1}^{T} D_{inv,k} \sim N(T\mu, T\sigma^2)$$

probability that Dpath exceeds a given timing constraint, which can be calculated with the CDF of a normal distribution. Although this simplified model trades off accuracy, it simplifies error rate prediction, allowing design optimization in a voltage-overscaled setting. If μ and σ are known, distributions of the various pipelinestages and modules in a system are estimated with this simple equation. Furthermore, timing slack redistribution or altering of pipeline depth may be performed without costly iterative circuit simulations.

C. Model Verification With Pipelined Baugh-Wooley Multipliers

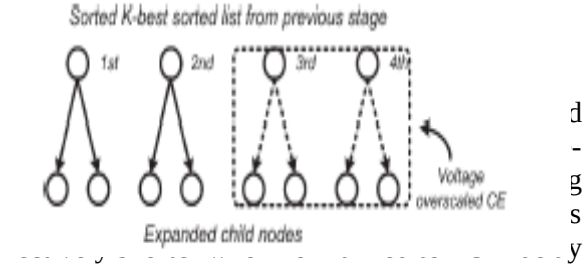
This section investigates the accuracy of the simple model described above in larger digital modules. Multipliers are used as a test case since they are one of the key components in DSP systems and are usually significantly larger than adders. In sub threshold design, leakage energy consumption can be suppressed by pipelining, allowing pipelined multipliers to un-pipelined multipliers .However, additional pipeline registers incur switching power overhead, and there exist an optimal number of pipeline stages that gives the lowest energy per operation. This section uses the proposed model to find an optimal pipeline depth for a voltage-overscaled multiplier. We implemented pipelined Baugh-Wooley multipliers with various pipeline depths and fabricated both un-pipelined and five-stage pipelined versions in a 65-nm CMOS technology (die photo in Fig. 4). Specific details on the design of an FFT accelerator using these Baugh-Wooley multipliers can be found in. The prediction model iscompared with both simulation and silicon measurement results. To validate the proposed delay model, we first determine whether the actual error rate distribution follows the Gaussian assumption.We performed measurements to obtain the error rate of the unpipelined multiplier with the result shown in Fig. 5. The plot indicates the linear relationship between the clock period and error rate quantile, indicating that the proposed model reflects the actual error probability reasonably well.





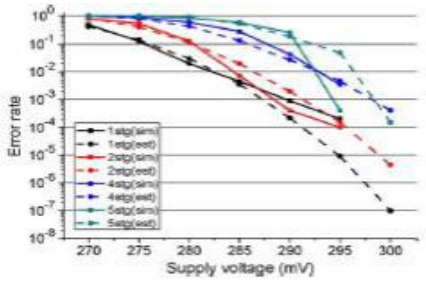
Given path delay ormal distrib lined multip N(μ_, σ_2) v based ition, pipelining may increase the number of first-order critical paths, which should be included in the error rate calculation. Hence, the error rate of an M-stage pipelined multiplier at a given clock period T is given by

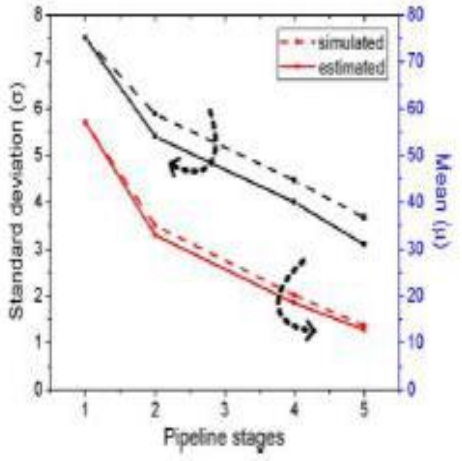
$$\mu' = \mu/M, \quad \sigma' = \sigma/\sqrt{M}$$



at the new voltage. Therefore, the effect of voltage scaling is directly reflected by T in this equation.

To measure the quality of the Gaussian-based predictionmodel, we implemented both un-pipelined and pipelined multipliers with pipeline depths of two, four, and five stages. The exact values of μ and σ for each multiplier are chosen by fitting (6) to simulation results, as seen in Fig. 6.We can also calculate the expected values of μ and σ by using (5), where K is defined as the ratio of worst-case logic-only critical path delays between the different pipeline implementations. This definition of K, rather than directly using the number of pipeline stages, compensates for sequential overhead and stage delay mismatch. Results from simulation and (5) are proposed model successfully predicts μ and σ with errors under 8.4% and 15.5%, respectively. Fig. 7 show measurement results of un-pipelined and fivestage pipelined multipliers along with simulated values and model predictions. Since the model calculates the probabilistic delay over the input vector space, there exists a lower limit of error probability. For example, if the worst-case critical path of a 16-bit RCA is activated with only one set of input vectors, the lowest possible error rate is 2−32. As the multiplier is pipelined more deeply and more critical paths are included, the probability of critical path activation increases significantly, raising the lower limit on the error rate. In Fig. 7, the error rate of the five-stage pipelined multiplier drops very quickly above 300 mV since the critical paths of Carry Save Adder trees are not activated, and timing errors only occur in the carry skip adder with slightly longer critical pathdelay. At 305mV,both multipliers are error free.





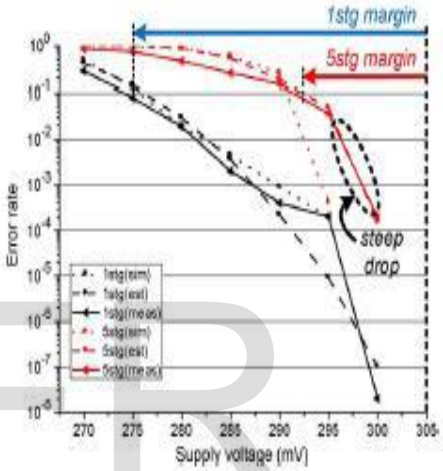


Fig: 6, 7, 8

Fig. 7 also shows the margins for VOS with a maximum error rate tolerance of 0.1. As also seen in Fig. 6, this margin reduces with increasing pipeline depth. A detailed analysis of the tradeoff between error rate and energy efficiency is given in Fig.8. Although a four-stage pipelined multiplier is the most energy efficient at error-free operation, a two-stage pipelined multiplier shows comparable efficiency at error rates of 10−3 ∼ 10−4 due to the steeper increase in error rates for deeply pipelined systems.

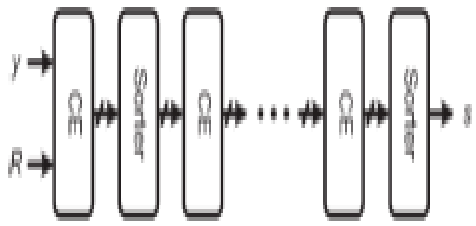


Fig: 10

4.CASESTUDY:ERROR-RESILIENTK-BEST DECODER

For modern and next-generation communication standards, the multiple-input and multiple-output technique is a key feature. The K-best decoder has been proposed for improved throughput and lower hardware cost with only a small SNR degradation. It accumulates the Euclidean distance from received symbols to candidate symbols and selects a fixed number of candidates with the minimum distances at each search stage. The K-best decoder consists of computational elements (CE) and sorters. The CE calculates the Euclidean distance to the most promising child nodes of the candidate lists from the previous stage and sends this information to the sorters. The sorters then order the child nodes based on distance and builds a new K-best candidate list.Fig.9 shows the K-best decoder architecture with received signal y, channel information R, and output symbol s. The K-best candidate list from the previous stage is already sorted, and the lower nodes on the list are less likely to survive until the last stage. This implies that even if an error occurs during node expansion of lower nodes, it will only have a minor impact on overall decoding performance. Therefore, if two separate CEs are used for the upper half and lower half nodes on the list, we can obtain error tolerance by employing a voltage-over scaled CE for the latter one, as seen in Fig. 10. A simple circuit-based timing error detection only scheme, such as in [3] and [4], can detect errors and set the calculated distance to infinity to remove erroneous results.SNR degradation due to voltage- overscaled CE in a practical K-best decoder design with K=10 and l=3 is described in Fig. 10, showing that the proposed error-resilient K-best decoder can tolerate a calculation error rate of a CE module up to 0.3 with SNR degradation under 0.4 dB at BER=10−3.The described CE module was synthesized in a 65-nm CMOS technology. The critical path consists of two adders and a multiplier in a series. Fig. 10 shows two pipelined critical paths with different a number of stages, while the remaining

5. CONCLUSION

We investigated the effect of VOS and proposed a framework for a voltage-over scaled DSP system design in the ultra-low voltage regime. We found that the error rate can be modeled using probabilistic critical path delays. This observation was generalized in the form of a Gaussian-distributed critical path delay model. This model enables rapid design decisions with reasonable accuracy and was verified against silicon pipelined Baugh-Wooley multipliers. An error-tolerant K-best decoder architecture that can tolerate CE error rates as high as 0.3 with SNR degradation < 0.4 dB was presented as a case study of a more complex system. We identified the optimal pipeline an error-resilient K-best decoder and showed that VOS enables 22.5% energy savings.

REFERENCES

[1] B. Zhai, S. Pant, L. Nazhandali, S. Hanson, J. Olson, A. Reeves, M. Minuth, R. Helfand, T. Austin, D. Sylvester, and D. Blaauw, “Energyefficient subthreshold processor design,” IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 17, no. 8, pp. 1127–1137, Aug. 2009.

[2] A. Wang, A. P. Chandrakasan, and S. V. Kosonocky,“Optimal supply and threshold scaling for subthreshold CMOS circuits,” in Proc. IEEE Comp. Soc. Annu. Symp. VLSI, 2002, pp. 5–9.

[3] D. Blaauw, S. Kalaiselvan, K. Lai, W.-H. Ma, S. Pant, C.Tokunaga, S. Das, and D. Bull, “Razor II: In situ error detection and orrection for PVT and SER tolerance,” in Proc. IEEE SSCC Tech. Dig., 2008, pp. 400–622. Wilkerson, S.-L.“Energy-efficient.

[4] K. A. Bowman, J. W. Tschanz, N. S. Kim, J. C. Lee, C. B. Lu, T. Karnik, and V. K. De, and metastabilityimmune timing-error detection and instruction-replay-based recovery circuits for dynamic-variation tolerance,” in Proc. IEEE ISSCC Tech. Dig., 2008, pp. 402–623.

[5] G. Karakonstantis, C. Roth, C. Berkeser, and A. Burg,On the exploitation of the inherent error resilience of wireless systems under unreliable silicon,” in Proc. IEEE 49th DAC, 2012, pp. 510–515.

[6] R. Liu and K. K. Parhi, “Power in frequency-selective FIR filters under voltage verscaling,”IEEE J. Emerging Sel. Top. Circuits Syst., vol. no. 3, pp.343–356, Sep. 2011.

[7] Y. Liu and T. Zhang,“On the selection of arithmetic unit design of energy-efficient processors,” IEEE processing,”in Proc.ACM/IEEE ISLPED, 2007, pp.250–255.

[8] R. A. Abdallah and N. R. Shanbhag, “Minimum-energy peration via error resiliency,” IEEE Embedded Syst. Lett.,vol. 2, no. 4, pp. 115–118, Dec. 2010.

[9] R. A. Abdallah, Y.-H. Lee, and N. R. Shanbhag, “Timing error statistics for energy-efficient robust DSP systems,” in Proc. DATE, 2011, pp. 1–4.

[10] J. Chen and J. Hu, “Energy-efficient digital signal processing viavoltageoverscaling-based residue number system,” IEEE Trans. Very Large Scale Integr. (VLSI) Syst.,to be published.

[11] A. B. Kahng, S. Kang, R. Kumar, and J.Sartori, “Recovery-driven design: Exploiting error resiliencein

Trans. Comput.-Aided Design Integr. Circuits Syst., vol. 31, no. 3, pp. 404–417, Mar. 2012.
[12] L.Wan and D. Chen, “Analysis of circuit dynamic behaviorwith timed ternary decision diagram,”in Proc.IEEE/ACM ICCAD, 2010, pp. 516–523.
[13] Y. Liu, T. Zhang, and J. Hu, “Design of voltage overscaled low-power trellis decoders in presence of process variations,” IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 17, no. 3, pp. 439–443, Mar. 2008. , Mar. 2006.