Impact of Delay Propagation on NTV PCMOS Design

L. Oudshoorn, G.A. Gillani, A.B.J. Kokkeler
Faculty of Electrical Engineering Mathematics and Computer Science, University of Twente, Netherlands

***Abstract*— While modern energy efficient low voltage designs focus on near-threshold voltage (NTV) operation for general purpose computing and exploiting an application's intrinsic error resilience by deploying probabilistic-CMOS (PCMOS) circuits for application specific computing, our results emphasize the impact of delay on PCMOS bits at NTV and lower voltage operations. This delay is very important to be considered while modeling PCMOS systems, as it propagates and has crucial effects on the most significant bits of the computations.**

I. INTRODUCTION

FUTURE low voltage noise dominated designs render probabilistic behavior of CMOS. This is acceptable as far as applications' intrinsic error resilience allows quantified inaccuracy in results to save energy consumption, such as in applications like audio/video processing and sky image formation in radio astronomy. This introduces the trade-off between energy consumption (E) and probability of correctness (p) that provides an opportunity for inexact computing to attain higher energy efficiency.

To understand why probabilistic behavior can lead to fundamentally lower energy usage, the process of switching can be analyzed from a thermodynamic perspective. Palem [1] has shown the energy gain of PCMOS as *kTln(1/p)*; where *p* is the probability of correctness, *T* is the temperature and *k* is the Boltzmann constant. Analytical discussion in [2] suggests to model a probabilistic switch, for instance an inverter, with a noise coupled output as,

$$p = P\left(X \le \frac{V_{dd}}{2}\right) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{V_{dd}}{2\sqrt{2}\sigma}\right) \quad (1)$$

where σ is the noise RMS. We will compare (1) with our simulation results in section II to discuss the relation in behavior of E-p curves at NTV and lower voltages.

Moreover, it is important to calculate the error propagation within the probabilitic system to find the total error which is bounded by the application's intrinisic resilience. In case of a probabilistic ripple carry adder, the probability of correctness of the sum output of stage $i+1$ depends on the probability of correctness of the adder block itself and also on the probability that its input carry from stage i is correct. Keeping in view this effect, M. Lau [3] calculated the propagation error within a 4-bit ripple carry adder for each sum and carry output. We further derive for the probability of correctness of sum outputs, i.e. the probabilty that the probabilistically calculated sum (s') at any stage (i) equals its deterministic counterpart (s) as,

$$P(s'_{i+1} = s_{i+1}) = \frac{1}{2} + \left(p^s_{i+1} - \frac{1}{2}\right) \times \left[\prod_{j=1}^{i}\left(p^c_j - \frac{1}{2}\right) + \sum_{k=1}^{i}\prod_{l=k}^{i}\left(p^c_l - \frac{1}{2}\right)\right] \quad (2)$$

Where p^c and p^s are the probabilities of correct outputs for carry and sum respectively. In section III, we will show from our simulation results, the impact of delay propagation in addition to error propagation as modeled in (2) and compare them to emphasize the importance of considering delay propagation in the PCMOS system design.

Based on the fact that modern digital design targets NTV region for optimal supply voltage [4], we are specifically interested in the energy consumption vs probability of correctness relation, i.e. E-p curves, within the NTV circuit operation. We have investigated the impact of variations in frequency of operation and noise levels on the E-p curves and found a different behavior between the analytical models and our simulation results in the NTV and lower levels of supply voltage.

II. SIMULATION SETUP AND RESULTING E-p CURVES

We have used the umc65 library in Cadence IC for the simulations. Our approach is to simulate 65nm technology with increased intrinsic noise due to channel resistance to represent the much smaller future transistors. A parameter called `noise scale' is used to amplify the noise levels. The points on the E-p curves are estimated by simulating many bit periods for a supply voltage setting and counting the number of correct and incorrect samples. A long transient simulation is performed using Cadence, which can simulate time domain noise. The results are then exported to Matlab where the sampling and processing of the data is done.

The simulation in Cadence is performed using a default set of parameters with variably increased noise amplification scales as shown in Table 1. In order to get all the points for the E-p curve, a parameter sweep is performed for the V_{dd} parameter. To keep the simulation time acceptable, the number of points (the supply voltage step size) for the E-p curves is kept relatively low. Although this results in a less smooth curve, it gives a reasonable representation of facts.

Table 1: Simulation Parameters.

Parameters	Value
MOS type	Umc65ll N/P_12_llrvt
NMOS gate length	60nm
NMOS gate width	80nm
PMOS gate length	60nm
PMOS gate width	160nm
V_{dd}	Range: 0-2V, 10mV steps
V_{in}	Alternating between 0V and V_{dd}
C_{out}	10fF
Temperature	27ºC
Noise amplification	50x/100x/200x/400x
Bit duration	1ns
# Bits simulated	800

The CMOS inverter circuit has been simulated to show the variations in E-p curves at low voltage levels. In our simulations, the noise can be scaled by a factor, which multiplies all generated noise by the chosen amount. Simulated E-p curves for noise scaled by a factor of 50 to 400 times are shown in Figure 1. Though noise scaling factors of over 100 are not realistic for the contemporary CMOS feature sizes, these numbers are chosen in order to better show the qualitative influence of noise on the E-p curves. The dotted line in Figure 1 represents the theoretical performance for a certain noise standard deviation (here we assume 100mV) according to (1). A large difference between the predicted shape and the simulation results is the sudden drop around the threshold voltage (≈0.5V) while lowering the supply voltage. This suggests that the inverter makes errors that are caused by malfunctioning rather than output misinterpretation due to noise. This is because the channel conductance of the inverter quickly becomes lower at low voltages, causing the output capacitance to charge (or discharge) slower. At a certain point the supply voltage becomes

too low such that the output capacitance is not charged before the sample moment, resulting in abrupt decrease in probability of correctness.

Simulations with various operating frequencies also demonstrate the delay of the circuit. Interestingly, the model used in (1) shown as dotted line in Figure 2, proves to be the theoretical maximum for the E-p tradeoff, which is approached by the simulated curves with the decrease in frequency of operation. To plot the dotted line, the standard deviation (σ) of the samples has been calculated for each supply voltage setting, which is assumed to be caused completely by the noise in the system. The theoretical maximum performance of a system in the presence of the measured amount of noise is calculated by filling in the supply voltage and noise standard deviation in the model specified by (1).

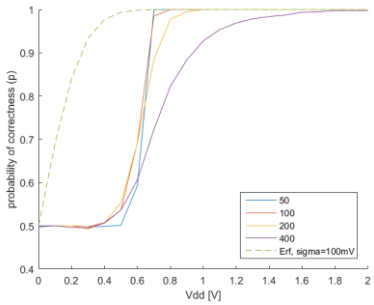

Fig. 1. CMOS inverter simulations for various noise scales.

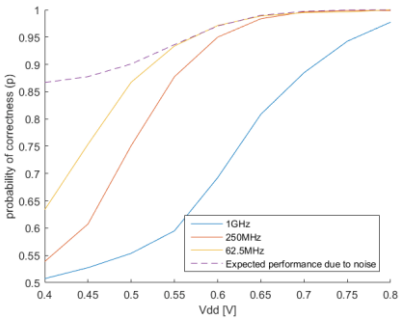

Fig. 2. CMOS inverter simulations for various operating frequencies.

III. DELAY PROPAGATION IN PCMOS SYSTEMS

We further investigated the influence that connected probabilistic building blocks have on each other. We simulated the 4-bit ripple carry adder comprised of 4 full adders in Cadence IC with the same assumptions as that of the inverter. Figures 3 and 4 present the simulated E-p curves for the carry and sum outputs respectively for the 4-bit ripple carry adder along with calculated ones according to (2). Theoretically, the outputs of stages 2, 3 and 4 are expected to be almost equal to that of stage 1, but in our simulations they are worse.

The theoretical curves are based on the assumption that the propagation of error is only due to probability of correctness metric. However, the delayed correct outputs of stage *i* can make the probability of correctness worse for stage *i+1* than calculated by (2). Therefore, a logical explanation for the theory and simulation not being equal is the delay propagation. At the start of a clock, the calculations are started using whatever value is present at the input from the previous calculation. There is a 50% chance that the next input will be different. However, the new value will not be available immediately. Therefore, the calculation may be underway when a new value settles on an input. The calculation of the output then starts again, but with less time left to complete it before the clock cycle ends. Unfortunately, this problem stacks for additional stages resulting in the most significant sum output to fail first.

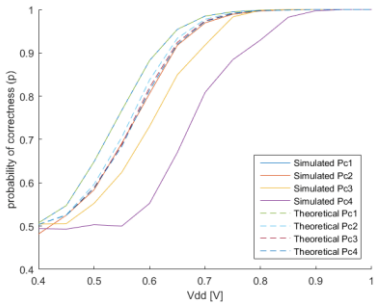

Fig. 3. Simulated and theoretical E-p curves for carry outputs

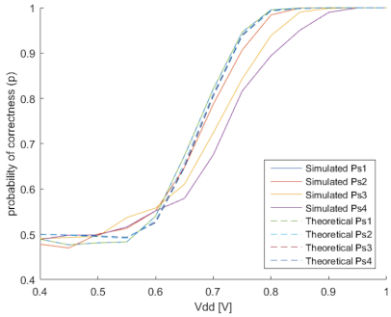

Fig. 4. Simulated and theoretical E-p curves for sum outputs

IV. CONCLUSIONS

We have simulated an inverter and a 4-bit ripple carry adder in Cadence that showed the shortcomings of current analytical models for the probability of correctness metric at near-threshold voltage (NTV) and lower supplies. We further investigated the impact of delay propagation in a digital system composed of probabilistic building blocks, which provides a clear insight of timing delay affecting the higher significant computational bits more than its lower significant counterparts and hence contributing considerably to the total error.

ACKNOWLEDGMENTS

This work was conducted in the context of the ASTRON and IBM joint project, DOME, funded by the Netherlands Organization for Scientific Research (NWO), the Dutch Ministry of EL&I, and the Province of Drenthe. Furthermore, we are thankful to dr. ir. A.J. Annema, ICD group, University of Twente, for simulation support.

REFERENCES

[1] KV Palem. “Energy aware computing through probabilistic switching.” Technical Report GIT-CC-03-16, Georgia Institute of Technology, 2003.

[2] Cheemalavagu, Suresh, et al. “A probabilistic CMOS switch and its realization by exploiting noise.” IFIP International Conference on VLSI. 2005.

[3] Lau, Mark SK, et al. “Modeling of probabilistic ripple-carry adders.” Electronic Design, Test and Application, 2010. DELTA'10. Fifth IEEE International Symposium on. IEEE, 2010.

[4] Pinckney, Nathaniel, David Blaauw, and Dennis Sylvester. “Low-Power Near-Threshold Design: Techniques to Improve Energy Efficiency.” Solid-State Circuits Magazine, IEEE 7.2 (2015): 49-57.