Ultra-Low Energy Computing with Noise: Energy-Performance-Probability Trade-offs*

Pinar Korkmaz Bilge E. S. Akgul Krishna V. Palem
Center for Research on Embedded Systems and Technology
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332
{korkmazp, bilge, palem}@ece.gatech.edu

Abstract

Noise susceptibility and power density have become two limiting factors to CMOS *technology scaling. As a solution to these challenges, probabilistic* CMOS *(*PCMOS*) based computing has been proposed.* PCMOS *devices are inherently probabilistic devices that compute correctly with a probability p. This paper investigates the trade-offs between the energy, performance and probability of correctness (p) of a* PCMOS *inverter. Using simple analytical models of energy, delay and p of a* PCMOS *inverter, the optimum energy delay product (*EDP*) value for given probability and performance constraints is found. The analytical models are validated using circuit simulations for a* PCMOS *inverter designed in a* $0.13\mu m$ *process. The results show that operating the* PCMOS *inverter at lower supply voltages is more preferable in terms of minimizing* EDP*. Our analysis is useful in optimal (in terms of* EDP*) circuit design for satisfying application requirements in terms of performance and probability of correctness. An analysis of the impacts of the variations in the temperature and the threshold voltage on the optimal* EDP *values is also included in the paper.*

1 Introduction

As CMOS technology scales down into the nano-meter region, significant challenges to sustaining Moore's law have emerged. Two of these challenges are achieving noise immunity (see Shepard [20], Natori and Sano [14]) and low-energy consumption (see [9, 12]). The conventional approaches to overcome these challenges encountered in the semiconductor roadmap view noise as an impediment to scaling (see Kish [10], Sano [19], Meindl [12]). As a paradigm shift from the conventional approaches, we have innovated PCMOS based computing in [4, 5] and [16], wherein noise is viewed as a resource rather than as an impediment for realizing ultra low-energy computing in the context of probabilistic applications. In [5] and [11], we characterized the energy consumed per switching step and the associated probability of correctness for a PCMOS inverter. It was also shown that PCMOS characteristics can be exploited at the application level for energy and performance benefits [4], wherein energy and performance benefits offered by PCMOS are quantified for a range of probabilistic applications.

In this work, we extend our characterization of a PCMOS inverter to include a succinct analysis of design trade-offs associated with its speed (or performance), energy and p. The characterization is achieved by using simple analytical models for energy, propagation delay, and p. In addition, we performed circuit simulations using BSIM3 models to verify our analytical model. In this paper, differing from our previous work [5, 11], we also consider leakage energy (in addition to the switching energy) of a PCMOS inverter, since the leakage energy is significant [17] especially for smaller feature sizes and for designs with low threshold voltages.

Lowering the supply voltage decreases the energy consumption, but also decreases p, which might be undesirable (depending on the value of p required by the application). Decreasing the supply voltage also decreases the switching speed of the circuit. Therefore, to meet the performance requirement demanded by the application, the threshold voltage should also be lowered. However, in this case, the static energy dissipation increases due to the increased leakage currents. Therefore, to study the trade-offs between energy consumption, performance and p, the parameters that we vary are the supply voltage (V_{dd}) and the threshold voltage (V_{th}). We also vary the RMS value of the noise to study the trade-offs between p, energy consumption and performance. In [9], Hegde and Shanbhag presented information-theoretic lowerbounds on the energy consumption of noisy gates. Their work is similar to our work since they also investigated the optimum values of V_{dd} and V_{th} that minimize the energy consumed by noisy gates. However, in their work, the primary focus is on computing reliably in the presence of noise, while our focus is on investigating the trade-offs between p (which is an independent design parameter), performance and energy. We also find optimal values of V_{dd} and V_{th} that satisfy p and performance requirements of an application, and minimize the EDP of PCMOS gates.

Supply voltage and threshold voltage scaling have been extensively studied (see [1, 2, 7]) in both the strong inversion and the subthreshold regions. The impact of V_{dd} and

*This work is supported in part by DARPA under Seedling # F30602-02-2-0124.

V_{th} on the energy and performance can be captured through the energy-delay product (EDP) metric that is commonly used to show the trade-offs between the two. In our work, we use the EDP metric to show the trade-offs between the energy, performance and p, with the goal of finding the optimal V_{dd}-V_{th} operation region for a PCMOS inverter. In particular, given a noise RMS value, a range of values for p and a performance constraint, we find the V_{dd} and V_{th} values that minimizes EDP. In addition, we consider the sensitivity of our analysis with respect to the variations in both the temperature and the threshold voltage. We show that our optimal V_{dd} and V_{th} operating points can change due to these variations.

Section 2 describes the PCMOS inverter. In Section 3, we describe our optimization procedure for finding the optimal values of V_{dd} and V_{th} that minimize EDP under given constraints on performance and a range of p values. In Section 4, we describe the impacts of V_{th} and temperature variations on the optimal values of V_{dd} and V_{th}. Finally, in Section 5, we conclude the paper.

2 Characterization of the Probabilistic Behavior of a PCMOS Inverter

A CMOS inverter is a digital switch that executes the *inversion* function with one input and output. For a deterministic inverter, $Y(t_2) = \overline{X(t_1)}$ where Y and X denote the binary values of the output and the input of the inverter, respectively, t_2 denotes the point in time when the switching ends, and t_1 denotes the point in time when the switching starts. For a probabilistic inverter, on the other hand,

$$Y(t_2) = \begin{cases} \overline{X(t_1)} & \text{with probability} \quad p \ (1/2 < p < 1) \\ X(t_1) & \text{with probability} \quad 1-p \end{cases} \tag{1}$$

where p denotes the probability of correctness for such an inverter. The probability p results due to the noise destabilizing the inverter. In this paper, we consider the case when thermal noise coupled to the output of the inverter is destabilizing the inverter. A comprehensive characterization of the PCMOS inverters in case of different couplings of noise can be found in [11].

We established in our prior work [5, 11] that a PCMOS inverter exhibits an exponential relationship between its p and the energy it consumes per switching, E. In addition, we showed that the relationship between the noise RMS value and the switching energy E is quadratic. The characterization of p and E derived from analytical modeling of noise susceptible CMOS inverters, has been extensively studied and verified using HSPICE simulations and physical measurements [11].

3 Trade-offs Between Energy, Performance and Probability of Correctness of a PCMOS Inverter

In this section, we explore the resulting values of energy, performance, and p for a range of values of V_{dd}: $0.30 \leq V_{dd} \leq 1.4$ and a range of values of V_{th}: $0.12 < V_{th} < 0.33$ for a PCMOS inverter realized in a 0.13μm process.

We consider an interval of p values, such as $0.90 < p < 0.95$ as seen from Figure 1, for which the design is being optimized. This interval of p values correspond to the bit error rate of the PCMOS device—in our case, the inverter—being optimized. Such a range of p values could reflect (1) the hardware-level degree of reliability of the device and (2) the application-level error tolerance range—and hence the quality—expected to be satisfied.

The hardware-level reliability, captured by the range of p values, is of interest for error redundancy mechanisms, such as NAND multiplexing studied by Norman et. al [15]. In their multiplexing scheme, a device is replicated N times, and the output values are compared according to a threshold, $\delta = p \in (0.5, 1)$, such that if the number of 0s (or 1s) is greater than $N \cdot \delta$, the output is decided to be 0 (or 1), whereas if it is in the interval $(N \cdot \delta, N \cdot (1-\delta))$, the output is undecided. Therefore, the individual p of the devices can show variation. Such a scheme would imply that given a range of p values, such as $(0.7, 1)$ for example, corresponding to the variation of the p of the device, our aim would be to optimize the individual performance of the devices in terms of energy and speed while preserving that the optimum EDP point still corresponds to the p interval.

As for the application-level error tolerance, a wide range of applications from the digital signal processing or image processing as well as the networking domains require a reliability threshold, which in turn reflects the application-level quality. The digital signal or image processing domain of applications have a certain range of error tolerance, typically characterized by signal-to-noise ratio (SNR) or distortion [8], whereas for networking, the reliability measure of communication channels are characterized through bit error rate and packet loss rate [6].

3.1 Modeling Energy, Performance and Probability of Correctness of a PCMOS Inverter

This section presents the models we used for propagation delay, leakage energy, switching energy and probability of correctness of a PCMOS inverter.

The propagation delay (t_g) of an inverter in the subthreshold region is described by

$$t_g = \frac{K_s C_L V_{dd}}{I_o} e^{\frac{V_{dd}-V_{th}}{n\phi_t}} \tag{2}$$

where K_s and I_o are fitted parameters (obtained using circuit simulations performed in HSPICE). C_L is the capacitive load for the inverter.

We find the propagation delay of an inverter in the strong inversion region, using a simple α-power law model [18]

$$t_g = K \frac{V_{dd}}{(V_{dd}-V_{th})^{\alpha}} \tag{3}$$

where K is a parameter fitted using circuit simulations. α is the velocity saturation constant which is also fitted using circuit simulations.

In modeling the leakage energy, gate leakage, and other leakage components, such as pn-junction leakage and gate-induced drain leakage, are neglected. We consider only the subthreshold leakage component for simplicity. Based on the BSIM3 v3.2 [23] equation for leakage energy consumed per switching cycle (during t_g) is described by

$$E_L = V_{dd} I_{th} \left(1 - e^{-\frac{V_{DS}}{n\phi_t}}\right) \cdot e^{\frac{-V_{th}-V_{off}+dibl \cdot V_{DS}}{n\phi_t}} \cdot t_g L_{DP} \tag{4}$$

where I_{th} denotes the channel current when $V_{GS} = V_{th}$, n is the body effect coefficient, V_{off} is an empirically determined BSIM3 parameter, $dibl$ is the DIBL (Drain Induced Barrier Lowering) factor, ϕ_t is the thermal voltage $\frac{kT}{q}$ and L_{DP} denotes the logic depth. We use a value of 25 for L_{DP}. This value is estimated based on the logic depth for the implementations [4] of the hyperencryption and probabilistic cellular automata algorithms. The values of V_{off} and $dibl$ are derived from curve fitting based on circuit simulations.

The switching energy, total energy per switching cycle, and EDP are described by (5) to (7)

$$E_{SW} = aC_L{V_{dd}}^2 \tag{5}$$

$$E_T = E_{SW} + E_L \tag{6}$$

$$\text{EDP} = E_T t_g \tag{7}$$

where a denotes the activity factor. In this paper, we assume that a is 10%. This value of a is chosen based on the activity factor of the PCMOS inverters used in the implementation of probabilistic applications [4] such as probabilistic Bayesian inference, random neural networks, and probabilistic cellular automata. The probabilistic content (as a percentage of total number of operations) of these applications varies from 0.25% in the case of the Bayesian inference to 19.7% in the case of the randomized neural network.

The probability of correctness for a PCMOS inverter is found using

$$p = 0.5 + 0.5erf\left(\frac{V_{dd}}{2\sqrt{2}\text{RMS}}\right) \tag{8}$$

We note that RMS denotes the standard deviation of the thermal noise that is coupled to the output of the inverter. In modeling the thermal noise, we follow the approach of [21], where the noise source is assumed to be a random process characterized by a Gaussian distribution. The details of derivation of (8) can be found in [11].

3.2 Optimal V_{dd} and V_{th} Operating Points

In this section, we employ the performance and p constraints imposed by an application on a PCMOS inverter to derive the optimal V_{th} and V_{dd} operating points that minimize the EDP of the inverter, for a given RMS value of noise. Such an optimization can be useful for architectural blocks (which implement probabilistic applications) whose minimum operating frequency needs to be greater than f_{min}, and whose reliability needs to be in a range, say, p_{min} to p_{max}. We now present the specific minimization problem under consideration, and the algorithm we have developed to solve the problem.

3.2.1 EDP Minimization Problem

We use the EDP metric to show the trade-offs between energy, performance and p of a PCMOS inverter. The performance of the PCMOS inverter is measured in terms of its maximum switching frequency, denoted as f_g, and is equal to the reciprocal of t_g.

In this section, we show normalized EDP (NEDP) contours, each denoting the ratio of the minimum EDP to the EDP corresponding to specific values of V_{dd} and V_{th}. To find the minimum EDP, we first find the values of V_{dd} and

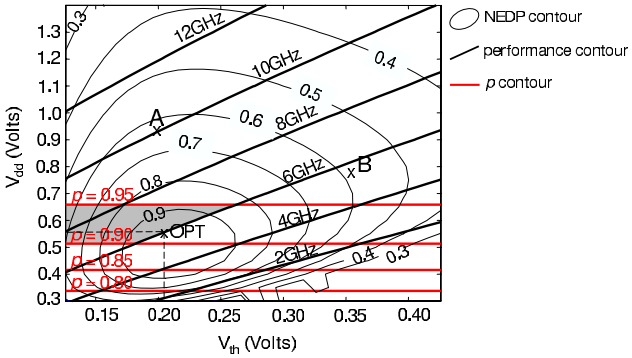

Figure 1. Constant NEDP, performance and p contours for a PCMOS inverter coupled with noise having an RMS value of 0.2V

V_{th} at which EDP is minimized. To find these values of V_{dd} and V_{th}, we differentiate (7) with respect to V_{dd} and V_{th} and equate the resulting equations to 0.

Figure 1 shows the NEDP, performance and p contours for a PCMOS inverter coupled with noise having an RMS value of 0.2V. In Figure 1 the rounded curves are contours of constant NEDP, the horizontal lines are contours of constant p, and the sloped lines are the contours of constant frequency (or performance). It is seen from the figure that, NEDP is high at lower values of V_{dd} and V_{th}. However, for very low values of V_{th}, NEDP becomes smaller due to the increased leakage energy. Figure 1 also shows that p increases as V_{dd} increases. We note that, the NEDP curves have small kinks near the border of subthreshold region, which are due to the discontinuity of the analytical model (equations (2) and (3)) at the boundary of subthreshold region. It is also seen from the figure that the higher the value of V_{dd} with respect to V_{th}, the higher the performance.

Given these trade-offs, our objective is to find the optimal V_{dd} and V_{th} operating points that minimize the EDP within the given constraints. The problem is stated as follows.

Minimize:

$$\text{EDP} = V_{dd}I_{th}\left(1 - e^{-\frac{V_{DS}}{n\phi_t}}\right) \cdot e^{\frac{-V_{th}-V_{off}+dibl\cdot V_{dd}}{n\phi_t}} {t_g}^2 L_{DP} + aC_L{V_{dd}}^2 t_g \tag{9}$$

subject to:

$$f_g \geq f_{min} \tag{10}$$

$$p_{min} \leq p \leq p_{max} \tag{11}$$

where t_g is the propagation delay of the inverter given by (2) or (3) depending on the operation region of the inverter.

The solution to this problem is found using a two-dimensional search algorithm whose pseudo-code is shown in Figure 2. As seen from Figure 2, for the given values of p_{min}, p_{max}, f_{min}, and noise RMS value:

1. We assign a sufficiently large value to EDP_{min}.
2. We increase p from p_{min} to p_{max} in sufficiently small steps. For each p:
   (a) We find the corresponding value of V_{dd} using (8).
   (b) Using (2) and (3), and given f_{min}, we find the maximum possible value of V_{th} (V_{thmax}).

1. Start: Input: p_{min}, p_{max}, f_{min} , noise RMS and $S >> 1$
2. $V_{thmin} = 0.125$; $V_{thstep} = V_{thmin} / S$; $p_{step} = p_{min} / S$;
3. i =0; j=0; $EDP_{min} = 1$;
4. p Loop: repeat
5. $i = i+1$; p(i) = $p_{min} + p_{step}$;
6. compute $V_{dd}(i)$ using (8);
7. compute V_{thmax} using (2) and (3);
8. $EDP_{min}(i) = 1$;
9. V_{th} Loop: repeat
10. $j = j+1$; $V_{th} = V_{thmin} + V_{thstep}$;
11. compute EDP using (9);
12. if EDP < $EDP_{min}(i)$
13. $EDP_{min}(i)$ = EDP;
14. $V_{thopt} = V_{th}(j)$;
15. $V_{ddopt} = V_{dd}(i)$;
16. until $V_{th}(j) > V_{thmax}$;
17. if $EDP_{min}(i) < EDP_{min}$
18. $EDP_{min} = EDP_{min}(i)$;
19. until p(i) > p_{max};
20. report EDP_{min}, V_{ddopt} and V_{thopt}

Figure 2. The pseudocode for the algorithm to obtain the optimal V_{dd} and V_{th} values that minimize EDP

(c) We assign a sufficiently large value to the minimum value of the EDP of this step ($EDP_{min}(i)$).

(d) We increase V_{th} from a given minimum value of V_{th} (V_{thmin}) to V_{thmax} in sufficiently small steps. For each V_{th}:

i. We compute EDP using (9). If this value of EDP is lower than $EDP_{min}(i)$, then the values of $EDP_{min}(i)$, V_{thopt}, and V_{ddopt} are updated as shown in steps 13, 14, and 15 of the pseudocode.

(e) If $EDP_{min}(i)$ is smaller than EDP_{min}, then we update the value of EDP_{min} as shown in line 18 of the pseudocode.

Referring to Figure 1, for example, if the performance constraint is set at 6GHz, the search algorithm searches for the optimal V_{dd} and V_{th} operating points in the region to the left of the 6GHz line. Furthermore, if p_{min} and p_{max} are 0.90 and 0.95, respectively, then the search is performed within the shaded area shown in the figure. The algorithm finds that for the optimal EDP point, the values of V_{dd} and V_{th} are 0.552V and 0.201V, respectively, as shown by the point OPT in Figure 1.

3.2.2 Simulation Results

In this section, we compare our analytical results with the simulation results for EDP, performance and p.

We performed circuit simulations in HSPICE using BSIM3 models for a CMOS inverter in a 0.13μm process to measure the inverter's static and dynamic energy consumption, propagation delay and p. We measured the static energy and the switching energy separately. We have assumed an activity factor of $a = 10\%$ in calculating the switching energy consumption.

In modeling the thermal noise that is coupled to the inverter, the noise source is assumed to be a random process characterized by a Gaussian distribution. The details of modeling the noise, the coupling of the noise and calculation of p in the circuit simulations can be found in [11].

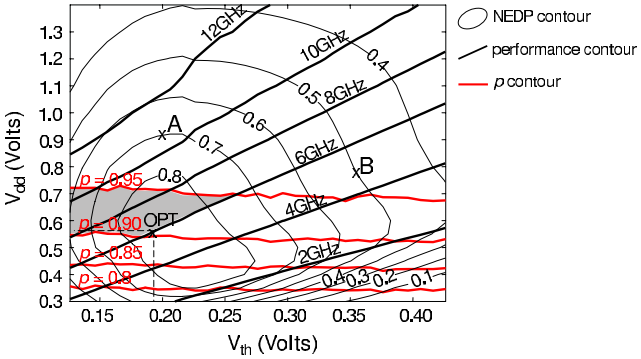

Figure 3. Constant NEDP, performance and p contours from circuit simulations

The results of the simulations are shown in Figure 3. When compared to Figure 1, the most striking difference is in the shape of the NEDP contours. The NEDP contours shown in Figure 3 are wider in the V_{dd} domain and narrower in the V_{th} domain compared to the NEDP contours in Figure 1. This difference results from the inaccuracy of the analytical model in estimating the propagation delay. The analytical delay model achieves an average error of 8.47%, but the standard deviation of the error is 9.47%. The analytical model overestimates the propagation delay for low values of V_{th}, and underestimates it for high values of V_{th}. Hence, as seen from Figures 1 and 3, at a fixed value of V_{th} and when V_{th} is small, the NEDP value from the analytical model is smaller than the NEDP value from the simulations. We can also see from Figures 1 and 3 that the analytical NEDP is higher than the simulated NEDP for higher values of V_{th}. This results from the underestimation of the propagation delay by the analytical model. For example, points denoted as A and B in both figures correspond to these two cases, where A represents the case when the analytical NEDP is lower, and B represents the case when the analytical NEDP is higher. As seen from Figures 1 and 3, analytical and simulation results for the performance contours are also deviating from each other due to the differences between the delay estimation in the analytical model and the simulations.

Comparing the p contours of Figures 1 and 3, we observe that the p contours found using simulations are traversing higher values of V_{dd}. This is caused by the fact that the transistors of the inverter used in simulations are not symmetrical, whereas the analytical model considers the case when the transistors are symmetrical. We have a more accurate model (see [11]) for the case when the transistors are not symmetrical. However, the more accurate model requires the midpoint voltage of the CMOS inverter, which we have not derived in the subthreshold region. Thus, we have chosen to use the model in (8) for simplicity in this paper. Furthermore, in Figure 3, the p contours are not exactly horizontal, but have a negative slope (which is very small in magnitude). This weak dependency of p on V_{th} is due to the dependency of p on the midpoint voltage of the inverter.

Comparing Figures 1 and 3, we see that the feasible region for the search example provided in Section 3.2.1 in case of simulations is slightly different from the feasible region in case of the analytical model. The optimal values of V_{dd} and V_{th} found from simulations are 0.55V and 0.196V

(as opposed to 0.552V and 0.201V). We note that to find the optimal operating points in case of simulations, we use a variant of the algorithm in Figure 2. We replace the steps for calculations of $V_{dd}(i)$, V_{thmax}, and EDP by search steps. The search step traverses the simulation results, and finds the closest values for $V_{dd}(i)$, V_{thmax}, and EDP.

As seen from Figures 1 and 3, the supply and threshold voltages for optimal EDP are closer to the probability contour $p = p_{min}$, that is, operating the PCMOS inverter at lower supply voltages is more preferable in terms of the EDP. However, the supply voltages can not be reduced further beyond the point where the performance constraint is satisfied.

4 Variations in Temperature and V_{th}

So far we have assumed that we have full control on the threshold voltages and the operating temperature. However in reality, the threshold voltage might change due to process variations and changes in the operating temperature. In addition, the chip temperature changes due to heat dissipation. Neglecting the coupling between the chip temperature and the power dissipation [1], in this section, we demonstrate the impact of the variations in the temperature and in the threshold voltage on the energy and performance of the PCMOS circuits in terms of the EDP contours derived in previous sections.

The operating temperature of a circuit can be anywhere between 25°C and 125°C. The threshold voltage at temperature T can be calculated [1] using

$$V_{thT} = V_{th} - k\,(T - T_{amb}) \tag{12}$$

where V_{thT} is the threshold voltage at temperature T, T_{amb} is the ambient temperature (25°C), and k is the threshold voltage temperature coefficient whose typical value for a 0.13μm process is 0.7mV/K [22].

The temperature also affects the RMS value of the noise, since we consider a thermal noise source. For simplicity, we assume that the noise source is a resistive noise source and therefore, we calculate the RMS value of noise at temperature T using

$$\text{RMS}_T = \text{RMS} \cdot \sqrt{\frac{T}{T_{amb}}} \tag{13}$$

Figure 4 shows the effect of increasing temperature from 25°C to 35°C. In the figure, the dashed contours correspond to the case when T is 35°C and the solid contours correspond to the case when T is 25°C. As seen from the figure, the NEDP and the performance contours are shifted to right when the temperature is increased. This results from the decrease in the threshold voltage due to the increase in the temperature. Furthermore, p contours are shifted higher in the V_{dd} domain, that is, to obtain the same value of p, a higher V_{dd} value is required at a higher temperature. Hence, at a fixed value of V_{dd}, p decreases as T increases. This decrease in p is due to the increased RMS value of noise due to the increased temperature. Due to these variations in NEDP, performance and p, the optimal values of V_{dd} and V_{th} also change. For example, with p_{min} and p_{max} values of 0.90 and 0.95, and performance constraint of 6GHz, the optimal values of V_{dd} and V_{th} are now 0.568V and 0.211V as opposed to the values of 0.552V and 0.201V found previously

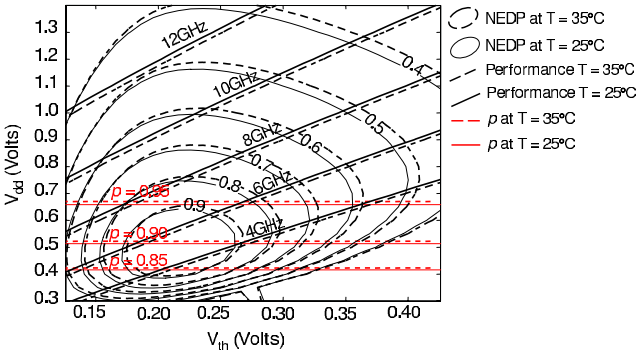

Figure 4. Constant NEDP, performance and p contours for a PCMOS inverter at temperatures $T = 25°$C and $T = 35°$C

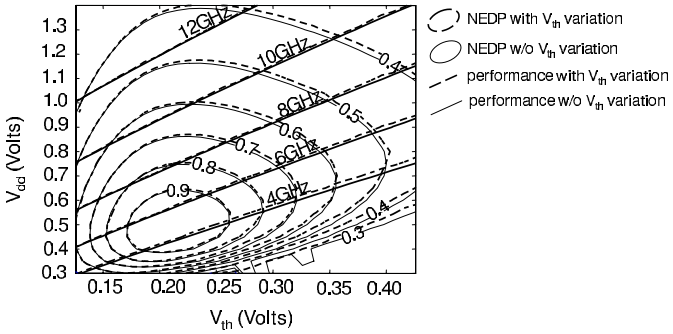

Figure 5. Constant NEDP and performance contours for a PCMOS inverter with V_{th} variations in case when mean values of EDP and performance are considered

in Section 3.2.1. As seen in this example, the change in optimal values of V_{dd} and V_{th} is small. However, if T is increased even further, we observe significant changes in the optimal values. For example, if T is 85°C, then the optimal V_{dd} and V_{th} values become 0.658V and 0.261V, corresponding to 19.2% and 22.9% difference when compared to the original values of 0.552V and 0.201V for $T = 25°$C.

Figure 5 depicts the effect of the variations in the threshold voltage. In the figure, the dashed contours correspond to the case when there is variation in the threshold voltage, and the solid contours correspond to the case when there is no variation in the threshold voltage. Empirical evidence suggests that the variation in V_{th} can be modeled by Gaussian distribution [13]. So, we model the threshold voltage variation as a Gaussian distribution. The mean value of the Gaussian distribution is 0 and its standard deviation σ is equal to 10% of the threshold voltage [3]. We only show the NEDP and performance contours since there is negligible effect of V_{th} on p. As seen from the figure, the NEDP contours are shifted upwards, that is, for a fixed value of V_{th}, to obtain the same value of NEDP, a higher V_{dd} value is required. This results from the fact that the change in EDP is larger when there is a positive change in V_{th} (eg. $V_{th}+\sigma$) compared to a negative change in V_{th} of the same magnitude (eg. $V_{th}-\sigma$). We note that, in Figure 5, we only show the mean values of NEDP and performance. As a result, the differences in NEDP and performance seem to be very small. Due to averaging,

the kinks in the NEDP contours have also disappeared. Furthermore, the optimal values of V_{dd} and V_{th} change slightly. For example, with p_{min} and p_{max} values of 0.90 and 0.95, and performance constraint of 6GHz, the optimal values of V_{dd} and V_{th} are now 0.562V and 0.206V corresponding to 1.81% and 2.49% difference compared to the optimal values of V_{dd} and V_{th} obtained in case when there is no V_{th} variation. If we consider a worst case scenario such as the mean plus one standard deviation, the difference in NEDP and performance becomes more significant.

We can conclude from the results of this section that EDP and performance are dependent on the variations in the T and V_{th}. Similarly, our V_{dd} and V_{th} values for optimal EDP are also dependent on these variations. Thus, an analysis for optimizing the EDP of a PCMOS inverter should consider the variations in T and V_{th}. Furthermore, we can conclude that threshold voltage control is necessary when the variations in T and V_{th} can not be modeled accurately.

5 Conclusions

In this paper, we have shown the design trade-offs between energy, performance and p of a PCMOS inverter using simple analytical models of energy, delay and p. We have also found the values of V_{dd} and V_{th} for optimal EDP under given constraints on p and performance. We have observed that operating the PCMOS inverter at lower supply voltages is more preferable in terms of minimizing EDP. We have also performed circuit simulations to validate our analytical models. From these simulations we have observed that the shapes of EDP surfaces and the location of the optimal EDP point are dependent on the models used for energy and delay. As an example, for a 0.13μm technology, given a minimum p requirement of 0.90, a maximum p requirement of 0.95 and a minimum frequency requirement of 6GHz, our analytical analysis yielded a supply voltage of 0.552V and threshold voltage of 0.201V, while our simulation results yielded a supply voltage of 0.55V and threshold voltage of 0.196V, which are reasonably close to each other. Our analysis can be helpful in circuit design for applications with a minimum performance requirement and a specific range of p.

We have also included an analysis of the impact of the variations in threshold voltage and temperature on EDP and performance contours as well as on optimal value of EDP. We have found that accurately estimating the variations in temperature and threshold voltage is important for accurately optimizing the EDP of a PCMOS inverter. This analysis can further be extended to include the variations in V_{dd}.

References

[1] A. Basu, S. C. Lin, V. Wason, A. Mehrotra, and K. Banerjee. Simultaneous optimization of supply and threshold voltages for low-power and high-performance circuits in the leakage dominant era. In *Proc. Design Automation Conf.*, pages 884–887, 2004.

[2] B. H. Calhoun, A. Wang, and A. Chandrakasan. Modeling and sizing for minimum energy operation in subthreshold circuits. *IEEE J. of Solid-State Circuits*, 40:1778–1786, 2005.

[3] Y. Cao, P. Gupta, A. B. Kahng, D. Sylvester, and J. Yang. Design sensitivities to variability: Extrapolations and assessments in nanometer VLSI. In *IEEE Intl. ASIC/SOC Conf.*, pages 411–415, 2002.

[4] L. N. Chakrapani, B. E. S. Akgul, S. Cheemalavagu, P. Korkmaz, K. V. Palem, and B. Seshasayee. Ultra efficient embedded SoC architectures based on probabilistic CMOS (PCMOS) technology. *To appear in Proc. of DATE Conference*, March 2006.

[5] S. Cheemalavagu, P. Korkmaz, K. V. Palem, B. E. S. Akgul, and L. N. Chakrapani. A probabilistic CMOS switch and its realization by exploiting noise. In *Proc. of the IFIP VLSI-SOC 2005*, Australia, October 2005.

[6] F. Gong and G. M. Parulkar. An application-oriented error control scheme for high-speed networks. *IEEE/ACM Trans. Netw.*, 4(5):669–683, 1996.

[7] R. Gonzalez, B. M. Gordon, and M. A. Horowitz. Supply and threshold voltage scaling for low power CMOS. *IEEE J. of Solid-State Circuits*, 32:1210–1216, August 1997.

[8] R. Hegde and N. R. Shanbhag. A low-power architecture for phase-splitting passband equalizer. In *IEEE Workshop on Signal Processing Systems, SIPS 97 - Design and Implementation*, pages 385–394, November 1997.

[9] R. Hegde and N. R. Shanbhag. Toward achieving energy efficiency in presence of deep submicron noise. *IEEE Trans. VLSI Syst.*, 8:379–391, August 2000.

[10] L. B. Kish. End of Moore's law: thermal (noise) death of integration in micro and nano electronics. *Physics Letters A*, 305:144–149, December 2002.

[11] P. Korkmaz, B. E. S. Akgul, and K. V. Palem. Characterizing the behaviour of a probabilistic CMOS switch through analytical models and its verification through simulations. Technical Report CREST-TR-05-08-01, Available at http://www.crest.gatech.edu/palempbitscurrent/.

[12] J. D. Meindl. Low power microelectronics: retrospect and prospect. *Proc. IEEE*, 83:619–635, April 1995.

[13] S. Narendra, D. Blaauw, A. Devgan, and F. Najm. Leakage issues in IC design: trends, estimation, and avoidance. In *Proc. of Intl. Conf. on Computer Aided Design*, November 2003.

[14] K. Natori and N. Sano. Scaling limit of digital circuits due to thermal noise. *J. of Applied Physics*, 83:5019–5024, May 1998.

[15] G. Norman, D. Parker, M. Kwiatkowska, and S. Shukla. Evaluating the reliability of nand multiplexing with prism. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 24(10):1629–1637, 2005.

[16] K. V. Palem. Energy aware computing through probabilistic switching: A study of limits. *IEEE Trans. Comput.*, pages 1123–1137, September 2005.

[17] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. *Proc. of the IEEE*, 91:305–327, 2003.

[18] T. Sakurai and A. R. Newton. Delay analysis of series-connected MOSFET circuits. *IEEE J. of Solid-State Circuits*, 26, 1991.

[19] N. Sano. Increasing importance of electronic thermal noise in sub-0.1mm Si-MOSFETs. *IEICE Trans. on Electronics*, E83-C:1203–1211, August 2000.

[20] K. L. Shepard. Design methodologies for noise in digital integrated circuits. In *Proc. Design Automation Conf.*, pages 94–99, June 1998.

[21] K.-U. Stein. Noise-induced error rate as a limiting factor for energy per operation in digital ICs. *IEEE J. Solid-State Circuits*, 12:527–530, Oct. 1977.

[22] Y. Taur and T. Ning. *Fundamentals of modern VLSI devices*. Cambridge Univ. Press, 1998.

[23] U.C. Berkeley BSIM Homepage. http://www-device.eecs.berkeley.edu/ bsim3/.