Modeling and Analysis of Leakage Power Considering Within-Die Process Variations

Ashish Srivastava Robert Bai David Blaauw Dennis Sylvester
University of Michigan, EECS Department, Ann Arbor, MI 48109
{ansrivas,baim,blaauw,dennis}@eecs.umich.edu

ABSTRACT

We describe the impact of process variation on leakage power for a 0.18µm CMOS technology. We show that variability, manifested in L_{drawn}, T_{ox}, and N_{sub}, can drastically affect the leakage current. We first present Monte Carlo-based simulation results for leakage current in various CMOS gates when the process parameters are varied both individually and concurrently. We then derive an analytical model to estimate the mean and standard deviation of the leakage current as a function of the process parameter distributions. We demonstrate that the results of the analytical model match well with Monte-Carlo simulations and also show the statistical mean leakage current is significantly different from the leakage predicted using a nominal case file.

1. INTRODUCTION

Supply voltage continues to be lowered in new process technologies to reduce dynamic power dissipation. However, in order not to degrade the speed of the device, the transistor threshold voltage (V_{th}) must be reduced commensurately. This has a drastic impact on the leakage power of the integrated circuit (IC) due to the exponential relationship between the threshold voltage and the leakage current (I_{off}) [1]. Large leakage currents shorten the battery life of mobile applications such as notebook computers, pagers, and PDAs that spend the majority of their lifetime in inactive mode. They can also severely degrade the noise immunity of dynamic logic circuits. In addition, short-channel effects such as V_{th}-roll off and drain-induced barrier lowering (DIBL) can impose serious obstacles to the scaling of V_{th}.

The computation of leakage current is complicated by the presence of process variation, which is developing as one of the foremost challenges in nanometer scale circuit design [2]. Pushing semiconductor technology and processing equipment to their limits results in appreciable uncertainty in key physical parameters such as gate oxide thickness and device channel length. In this context, performance has solely been defined as speed while power, and particularly static power consumption, has not been addressed at all. The exponential relationship between I_{off} and V_{th} can lead to enormous fluctuations in static power due to process variations. Previously, when static power was a negligible component of the total IC power budget, ignoring its variation was acceptable. Today, however, this leads to significant problems for two reasons: 1) Controlling V_{th} becomes more difficult in scaled technologies due to discrete dopant effects and V_{th} roll-off. Discrete dopant effects refer to the fact that in a very small channel there are only a small number of dopants present. The placement of these dopants and random fluctuations in their number, can lead to substantial changes in V_{th} from device to device. 2) Static power currently already accounts for ~20% of the total power budget in high-end microprocessors [3] and this number will likely increase with further reductions in V_{th}.

Furthermore, due to the exponential dependency of static power on V_{th}, statistical modeling based on corner models in this case becomes even less practical than in the modeling of delays. Relying on simple corner models would result in excessive guardbanding with the penalty being reductions in speed due to the use of higher than necessary threshold voltages. At the same time, neglecting the variability of static power consumption will surely result in underestimation, as the devices with smaller V_{th}'s will dominate the total leakage current budget. This will cause battery life in mobile applications to be substantially shorter than expected.

In summary, the combination of rising static power consumption and V_{th} controllability problems make the issue of statistical modeling of leakage current a pressing issue in future high-performance IC design. This paper provides the first work in this area and we focus on obtaining insight to the key physical parameters to model in I_{off} variability as well as deriving analytical expressions to circumvent the computational complexity of Monte Carlo analysis. In Section 2, we present HSPICE simulation results of leakage current for a single transistor and a 3-input NAND gate as we model three process parameters that have direct impact on Vth, namely L_{min}, T_{ox} and N_{sub} as Gaussian random variables. In Section 3 we analytically derive characteristics of the leakage current distribution given process parameter specifications. In Section 4 we compare the analytical model with the simulation results from Section 3. Section 5 provides conclusion and describes future work.

2. SENSITIVITY ANALYSIS

The primary goal of the simulations was to assess the extent of the leakage current variation as a result of process variations in gate length (L_{drawn}), gate oxide thickness (T_{ox}) and the channel dose (N_{sub}). The distribution of these parameters was assumed to be Gaussian; this should be a more reasonable assumption than dictating the threshold voltage of the devices to be Gaussian. The structural and doping parameters for the nominal device along with the amount of statistical variation used for each of the three parameters were in accordance with [4]. The gate oxide thickness and channel doping had a variance of 10%. Since variation in L_{eff} is most important, its effect was studied at both 10% and 20% variation to reflect both aggressive and realistic processes. The

This work was supported in part by the MARCO/DARPA Gigascale Silicon Research Center (http://www.gigascale.org) and the Semiconductor Research Corporation under contracts 2001-TJ-915 and 2001-HJ-959

analysis was carried out in two ways: We first varied the parameters one at a time, i.e., one of the parameters was varied while all other parameters were held constant. In the second stage we vary all parameters concurrently although each parameter is statistically independent of the others. The former analysis demonstrates the dependence of leakage on the particular parameter being varied, while the latter gives a more complete and accurate picture of the combined effects of within-die process variation on leakage current. For both types of analysis we used 10,000 HSPICE simulations. Each time the process parameters of interest take on a value from a normally-distributed probability-density function with the specified statistical mean and variance. We performed both sets of analyses for an inverter and a NAND stack. First, under the room temperature environment, the inverter was simulated for both possible input combinations. The results obtained for a low input (NMOS leakage) are shown in Table 1.

Table 1: Leakage variation of a NMOS/PMOS device with variation in process parameters at room temperature

Parameter Varied	3σ Var.	Mean Leakage (pA) N/P	Standard Dev. N/P	S.D./Mean N/P
None (Nominal)		42.4/26.4		
N_{ch}	10%	42.5/26.5	1.8/1.0	4.2/3.8 %
T_{ox}	10%	42.9/27.0	9.0/6.2	21/23 %
L_{drawn}	10%	44.1/32.0	9.6/22.0	21.8/68.8 %
L_{drawn}	20%	52.4/78	41.5/227	79.2/291 %
$N_{ch,}$ $T_{ox,}$ L_{drawn}	10,10, 10%	45.9/33.6	15.7/27.8	34.2/82.7 %
$N_{ch,}$ $T_{ox,}$ L_{drawn}	10,10, 20%	54.2/89	45.0/310	83.0/348 %

The effect of the variation in gate length has a severe effect on the leakage current, while the variation in channel dose has little impact. Table 1 shows the leakage results for a PMOS device. The results clearly show that the degradation of PMOS leakage current with variations in the gate length is much worse than an NMOS counterpart with the same degree of gate length variation. This arises since V_{th} roll-off due to short-channel effects for PMOS devices tends to be worse than that of NMOS devices [5]. Again in the case of PMOS transistor, the impact of the variation in the channel dose is found to be insignificant when compared to the effect of the variation in the gate length. Therefore, for a single inverter, we conclude that the leakage of the PMOS device would dictate the process control specifications because of the more pronounced susceptibility of PMOS transistors to process variation, particularly to gate length variability.

To look at the impact of variability in realistic operating conditions, the devices were re-simulated at an operating temperature of 100°C. For the NMOS case, the relative spread of leakage current (as quantified by the ratio of standard deviation to mean) is roughly the same at elevated temperature and room temperature. In PMOS devices, we find a substantial reduction in the relative spread of leakage. However, since the magnitude of leakage is much higher at 100°C as expected, the actual uncertainty in leakage has increased in this instance as well. As a result, designers of high-end ICs expected to run at high temperatures need to be particularly aware of variability in static power consumption. Figure 1 shows the distribution of the leakage current due to varying gate lengths. The distribution is skewed towards higher leakage values because of the exponential relationship from [1] and is further enhanced by a (1/L) term in I_{off}, as will be shown in Equation 6.2 of Section 4.

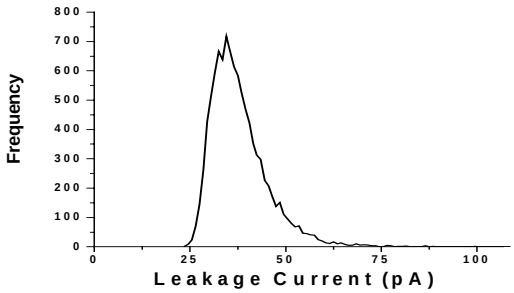

Figure 1. **Distribution of leakage current considering variability in gate length.**

The leakage current analysis for NAND/NOR stacks is inherently more complicated than for a single inverter because the net leakage effect is highly dependent on the applied input vectors. As shown in [6] when two leaking transistors are in series the leakage current is significantly reduced due to the "stack effect," which refers to the leakage reduction effect in a transistor stack when more than one transistor is turned off.

Table 2: **State-dependent leakage current of a 3-Input NAND**

Circuit State (ABC)	Nominal Leakage (pA)	Mean Leakage (pA)	Standard Deviation (pA)
000	5.71	6.17	1.89
001	8.86	9.79	4.06
010	8.95	9.88	4.07
100	11.6	12.5	3.8
101	38.9	48.1	80
110	41.3	50.2	74
011	42	57	78
111	80	256	695

The leakage depends on the number of off transistors in the stack, because the on transistors are treated as short circuits [4]. Considering the spatial proximity of series transistors within a gate, we assume the gate lengths to be perfectly correlated in this work. Table 2 shows the leakage current for a 3-input NAND gate for all eight possible states when L assumes a 20% variation (3σ) in gate length. As expected, due to the stack effect the mean leakage decreases significantly as we go from one leaking transistor to two leaking transistor in series, and when all three NMOS transistor in series are off, the leakage current is smallest due to the maximized stack effect. The case when all three PMOS transistors are off has the maximum leakage current since there are three leaking PMOS transistors in parallel. Finally, we note the difference in nominal and average leakage – the mean leakage over 10,000 instances of 3-input NANDs can be anywhere from 8% to 320% higher than the nominal leakage case due to the skewed distribution as shown in Figure 1.

3. ANALYTICAL APPROACH

The main disadvantage of Monte Carlo simulation its high computational effort which makes it impractical for analysis during the design process. Hence, we propose an analytical model that allows efficient computation of the mean and standard deviation of leakage current. We derive a model for computing the mean and standard deviation for a single transistor. We study the dependence of leakage on the three parameters: gate length, oxide thickness, and channel doping.

The subthreshold equation is expressed as:

$$I_{sub} = I_0 \exp((V_{gs} - Vth)/nV_T)(1-\exp(-V_{ds}/V_T)) \quad \textbf{(1)}$$

where V_T = KT/q and $I_0=\mu_0 C_{ox}(W_{eff}/L_{eff})V_T{}^2 e^{1.8}$. V_{th} is the threshold voltage of the device and is given by:

$$Vth = V_{fb} + \left|2\phi_p\right| + \frac{\lambda_b}{C_{ox}}\sqrt{2qN\varepsilon_s(\left|2\phi_p\right| + V_{sb})} - \lambda_d V_{ds} \quad \textbf{(2)}$$

where λ_d is the DIBL coefficient, and is expressed as in [7] which has been verified for accuracy down to L_g =0.1μm

$$\lambda_d = \left[\frac{L}{2.2\mu m^{-2}(T_{ox}+0.012\mu m)(W_{sd}+0.15\mu m)(X_j+2.9\mu m)}\right]^{-2.7} \quad \textbf{(3)}$$

and λ_b is the body-effect factor and is expressed as

$$\lambda_b = 1-\left(\sqrt{1+2W/X_j}-1\right)X_j/L \quad \textbf{(4)}$$

we now consider the case of a single inverter where the NMOS is turned off and $V_{gs} = 0$. From (1) we get

$$I_{sub} = I_0(1-\exp(-V_{dd}/V_T))\exp(-Vth/nV_T) \quad \textbf{(5)}$$

which shows that the leakage current is a function of the threshold voltage, drawn dimensions and gate oxide thickness. We then examine the three specific cases where we vary the gate length, the gate oxide thickness and the channel doping for the single inverter.

3.1 Case I: Variation in gate length

The change in threshold voltage due to change in gate length is obtained using $\frac{\Delta Vth}{\Delta L} = \frac{\partial Vth}{\partial \lambda d}\frac{d\lambda_d}{dL} + \frac{\partial Vth}{\partial \lambda_b}\frac{d\lambda_b}{dL}$

and the expression for threshold voltage (2), which gives

$$\Delta Vth = \left[\left(2.7V_{dd}\lambda_d + \frac{1}{C_{ox}}\sqrt{2qN\varepsilon_s(\left|2\phi_p\right| + V_{sb})}(1-\lambda_b)\right)\frac{1}{L}\right]\Delta L = K_l \Delta L \quad \textbf{(6.1)}$$

Having approximated the change in threshold voltage to be linear with the change in the process parameter, gives an exponential dependence of the leakage current on the variation of the process parameter. If we look at the process parameter as a random variable then the distribution of the exponential would be lognormal, but the dependence of I_0 on the process parameter skews the distribution of the leakage current further. Taking this into consideration (5) can be written as

$$I_{sub} = K(1/L)\exp(-(K_l/nV_T)L) \quad \textbf{(6.2)}$$

The mean of a random variable is defined as

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) dx,$$

if x is concentrated near its mean, E(g(x)) can be expressed in terms of the moments of x [8]. Since the variations of the process parameters are within the range of 10-20% of the mean value, we can assume the parameters to be concentrated near their mean. Applying Taylor series theorem to the above expression we can re-write it as follows,

$$E(g(x)) = \int_{-\infty}^{\infty}\left(g(\eta) + g'(\eta)(x-\eta) + ... + g^n(\eta)\frac{(x-\eta)^n}{n!}\right) f(x)dx,$$

where η is the mean of f(x), which gives,

$$E(g(x)) \; ; \; g(\eta) + g'(\eta)\mu_1 + g''(\eta)\mu_2 + ... + g^{(n)}(\eta)\frac{\mu_n}{n!}. \quad \textbf{(7)}$$

where the μ's are the central moments of f(x), which is Gaussian by construction. For our case $g(x) = e^{-Kx}/x$ and a third order approximation of g(x) should suffice. Also, since f(x) is Gaussian, $\mu_1 = \mu_3 = 0$. Using these conditions, we get,

$$E(g(x)) = \exp(-K\eta)\left(\frac{1}{\eta} + \frac{2\sigma^2}{\eta^3} + \frac{2K\sigma^2}{\eta^2} + \frac{K^2\sigma^2}{\eta}\right) \quad \textbf{(8.1)}$$

where σ is the standard deviation of f(x). Similarly,

$$E(g(x)^2) = \exp(-2K\eta)\left(\frac{1}{\eta^2} + \frac{6\sigma^2}{\eta^4} + \frac{8K\sigma^2}{\eta^3} + \frac{4K^2\sigma^2}{\eta^2}\right). \quad \textbf{(8.2)}$$

Using (8) and (6) we get the mean and the variance of the leakage current.

3.2 Case II: Variation in oxide thickness

Following the same approach as in the previous case, using (2) and

$$\frac{\Delta V_{th}}{\Delta T_{ox}} = \frac{\partial V_{th}}{\partial \lambda_d}\frac{d\lambda_d}{dT_{ox}} + \frac{\partial V_{th}}{\partial C_{ox}}\frac{dC_{ox}}{dT_{ox}}$$

we obtain,

$$\Delta Vth = \left[\frac{\lambda_b}{\varepsilon_{ox}}\sqrt{2qN\varepsilon_s(\left|2\phi_p\right| + V_{sb})} - \frac{2.7\lambda_d V_{dd}}{(T_{ox}+.012\mu m)}\right]\Delta T_{ox} = K_{T_{ox}}\Delta T_{ox} \quad \textbf{(9.1)}$$

Since $I_0 \propto (1/T_{ox})$, (5) can be expressed as

$$I_{sub} = K(1/T_{ox})\exp(-(K_{T_{ox}}/nV_T)T_{ox}) \quad \textbf{(9.2)}$$

The subthreshold current has similar dependence on the gate oxide thickness as on the gate length. Thus (8) holds and the mean and variance of the leakage current can be evaluated using (8) and (9).

3.3 Case III: Variation in channel doping

I_0 is independent of the channel doping concentration; hence we only need to look at the variation of the threshold voltage. Using (2) and $\phi_p = V_T \ln\left(\frac{N_{sub}}{N_i}\right)$,

$$\Delta V_{th} = \left[\frac{2V_T}{N_{sub}} + \frac{2q\lambda_b\varepsilon_s V_T(\ln N_{sub} - \ln N_i + 1)}{C_{ox}\sqrt{4qN_{sub}\varepsilon_s V_T(\ln N_{sub} - \ln N_i)}}\right]\Delta N_{sub} = K_{N_{sub}}\Delta N_{sub}$$

(10.1)

In this case (5) can be expressed as

$$I_{sub} = K \exp(-(K_{N_{sub}} / nV_T) N_{sub}), \quad \textbf{(10.2)}$$

so g(x) = e^{-Kx}. This has a lognormal distribution as discussed. From the above, we obtain:

$$E(g(x)) = \exp(-K\eta + \frac{k^2\sigma^2}{2}) \quad \textbf{(11.1)}$$

$$E(g(x)^2) = \exp(2(K\eta + K^2\sigma^2)) - \exp(2K\eta + K^2\sigma^2) \quad \textbf{(11.2)}$$

Now (10) and (11) can be used to obtain the mean and variance of the leakage current. The constant K in (6.2), (9.2), and (10.2) can be obtained either analytically or by using the nominal leakage value of the gate.

3.4 Simultaneous variation of multiple parameters

The problem of evaluating the leakage when all the parameters are varying simultaneously is simplified by our earlier approximation of linearizing the effect of the change in threshold voltage with the process parameters (see Equations 6.1, 9.1, and 10.1). Under these assumptions the expression of the subthreshold current takes the form

$$I_{sub} = \frac{K}{LT_{ox}} \exp\left(\frac{-K_L \Delta_L}{nV_T}\right) \exp\left(\frac{-K_T \Delta_{T_{ox}}}{nV_T}\right) \quad \textbf{(12)}$$

Given two random variables X and Y, if they are independent then the expectation of their product is the same as the product of their expectation. Using this fact, we can estimate the mean and standard deviation of the leakage current assuming the variation of the parameters to be independent of each other.

$$E(XY) = E(X)E(Y)$$

$$Var(XY) = E((XY - E(XY))^2) = E(X^2)E(Y^2) - (E(XY))^2$$

Where E(X) is the expected value of X alone and E(Y) is the expected value of Y alone. Hence the variance of the leakage current can be expressed more explicitly in terms of the variance when the process parameters are varying one at a time.

$$Var(XY) = (Var(X) + (E(X))^2)(Var(Y) + (E(Y))^2) - (E(X))^2 (E(Y))^2$$

4. RESULTS

In this section we compare the results from the sensitivity analysis and the analytical approach of the previous two sections. Table 3 shows the comparison for the inverter. The results show that there is generally good agreement between the experimental and analytical results. In particular, the analytical results for the mean match very well with the simulated values. The analytical expressions are expected to yield poor results for large variations (e.g. ≥20%) since the effect of the variation of the parameters on the threshold voltage has been assumed to be linear. A higher order expression for the change in the threshold voltage can be used for large variations, but would further complicate the mean and variance expressions.

5. CONCLUSIONS AND FUTURE WORK

We have shown that within-die process variation can have a significant impact on the leakage current for CMOS transistors. To quantify the effect, we have run Monte-Carlo simulations on a CMOS inverter. It is found empirically that the leakage current is exponentially proportional to the change in gate length and gate oxide thickness, but it is linear to the change in channel doping. We point out the significant difference between the average leakage over a large number of gates and the nominal leakage in a gate with typical process parameters.

Table 3: **Analytical (Anlyt) and experimental (Exp) statistical data comparison for a single device**

Parameter Varied	% variation of Parameter (3 σ)	Mean Leakage (pA) (Exp)	Mean leakage (pA) (Anlyt)	S.D.of leakage (pA) (Exp)	S.D.of leakage (pA) (Anlyt)
T_{ox}	10%	42.9	42.9	9.0	6.1
N_{sub}	10%	42.5	42.4	1.8	2.0
L_{drawn}	10%	44.1	44.5	9.6	12.9
L_{drawn}	20%	52.4	50.7	41.5	25
$N_{ch,}$ $T_{ox,}$ L_{drawn}	10,10,10%	45.9	45.0	15.7	14.0
$N_{ch,}$ $T_{ox,}$ L_{drawn}	10,10,20%	54.2	51.2	45	26

We also derived equations for computing the statistical mean and variance for the leakage current in a single inverter given the mean and variance of some typical process parameters. The results obtained via this analytical method agree well with those obtained experimentally.

In our future work, we intend to enhance the analytical model to be able to compute the statistical mean and variance for *n* stacked devices given the statistical data for the process parameters. We have done some preliminary research in this area and found that a simple model that we had used which included only a stack factor [7] did not produce good results. A set of more robust equations for computing the leakage current of stacked transistors is required.

6. REFERENCES

[1] C. Hu, "Device and technology impact on low power electronics," in *Low Power Design Methodologies*, ed. Jan Rabaey, Kluwer, pp. 21-35, 1996.

[2] International Technology Roadmap for Semiconductors, 2001.

[3] http://developer.intel.com/design/mobile/datashts

[4] P.M. Zeitzoff, A.F. Tasch, W.E. Moore, S.A. Khan, and D. Angelo, "Modeling of manufacturing sensitivity and statistically based process control requirements for a 0.18μm NMOS device," *Intl. Conference on Characterization and Metrology for ULSI Technology*, 1998.

[5] S. Tyagi et al, "A 130nm generation logic technology featuring 70nm transistors, dual Vt transistors and 6 layers of Cu interconnect," *Proc. IEDM*, pp. 567-570, 2000.

[6] Z. Chen, M. Johnson, L. Wei, and K. Roy, "Estimation of standby leakage power in CMOS circuits considering accurate modeling of transistor stacks," *Proc. ISLPED*, pp. 239-244, 1998.

[7] S. Narendra, D. Antoniadis, "Impact of using adaptive body bias to compensate die-to-die Vt variation on within-die Vt variation," *Proc. ISLPED*, pp. 229-232, 1999.

[8] A. Papoulis, "Probability, Random Variables, and Stochastic Processes," McGraw-Hill, Inc, New York, 1991

[9] R. Gu, M. Elmasry, "Power dissipation analysis and optimization of deep submicron CMOS digital circuits," *IEEE Journal of Solid-State Circuits*, v. 31, no. 5, pp. 707-713, May 1996.