5.4

A Probabilistic Framework to Estimate Full-Chip Subthreshold Leakage Power Distribution Considering Within-Die and Die-to-Die P-T-V Variations

Songqing Zhang, Vineet Wason and Kaustav Banerjee
Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106
{szhang, vwason, kaustav}@ece.ucsb.edu

ABSTRACT

This paper presents a probabilistic framework for full-chip estimation of subthreshold leakage power distribution considering both within-die and die-to-die variations in process (P), temperature (T) and supply voltage (V). The results obtained under this framework are compared to BSIM results and are found to be more accurate in comparison to those obtained from existing statistical models. Using this framework, a quantitative analysis of the relative sensitivities of subthreshold leakage to P-T-V variations has been presented. For the first time, the effects of die-to-die channel length and temperature variations on subthreshold leakage are studied in combination with all within-die variations. It has been shown that for accurate estimation of subthreshold leakage, it is important to consider die-to-die temperature variations which can significantly increase the leakage power due to electrothermal couplings between power and temperature. Furthermore, the full-chip leakage power distribution arising due to both within-die and die-to-die P-T-V is calculated, which is subsequently used to estimate the leakage constrained yield under the impact of these variations. The calculations show that the yield is significantly lowered under the impact of within-die and die-to-die process and temperature variations.

Categories and Subject Descriptors

B.7.1 [**Integrated Circuits**]: Types and Design Styles – *VLSI.*

General Terms: Performance, Design.

Keywords: Subthreshold leakage power distribution, yield estimation, process variations, die-to-die variations, within-die variations, electrothermal couplings.

1. INTRODUCTION

For sub-100 *nm* CMOS technologies, within-die and die-to-die variations in process (P), supply voltage (V) and temperature (T) are resulting in an increasing spread in the distribution of performance metrics such as delay, power, robustness and reliability. In particular, leakage power that constitutes an increasing component of total power has been reported to have 20X variations for a 180 *nm* CMOS technology [1]. Thus, designing with the worst-case leakage values may result in excessive guard-banding while underestimating the leakage might result in highly optimistic designs. Therefore, in the present scenario, probabilistic modeling is more meaningful in comparison to a deterministic analysis. Additionally, due to a 5X increase of total leakage power every generation [2], the design constraint based on leakage power may soon limit the yield. Therefore, it is critical to develop a probabilistic framework for accurately estimating full-chip subthreshold leakage power distribution under P-T-V variations which can be subsequently used to accurately estimate the yield. Furthermore, a quantitative analysis of the relative sensitivities of leakage to P-T-V variations is highly desirable so that relevant variations can be targeted to improve the yield.

Although, existing work [3] has successfully quantified the impact of parameter variations on performance, but there is no known work that has sufficiently described the combined effects of P-T-V variations on leakage power. Su et al. [4] estimated the full chip leakage considering uneven voltage drop and uneven temperature but it is not a probabilistic approach and hence can't be used for estimating yield. Recently, Rao et al. [5] studied the impact of channel length variations on subthreshold leakage, but their analysis was based on an empirical relationship between leakage and channel length which cannot be easily extended to other variations such as oxide thickness or temperature variations since that would first require an empirical and invertible relationship between leakage and oxide thickness or temperature. Although, the work presented in [6]-[8] develop statistical models to estimate the leakage under variations, but they do not account for the combined effects of within-die and die-to-die P-T-V variations. For instance, Narendra et al. [7] analyzed the impact of only within-die V_{th} variation on subthreshold leakage whereas Mukhopadhyay et al. [6] analyzed the sensitivity of various leakage components to within-die process and voltage variations only. Also, Srivastava et al. [8] analyzed the impact of only within-die process variations. Moreover, due to the approximations involved, these analyses are inaccurate when compared to BSIM models as will be discussed later in the paper. Also, these models cannot be used to estimate the yield since they do not provide the probability distribution function of the leakage power.

In this paper, a probabilistic framework to simultaneously analyze the impact of both within-die and die-to-die P-T-V variations on subthreshold leakage power for sub-100 nm CMOS technologies has been introduced. We focus on subthreshold leakage since it is the most dominant component of total leakage in high-performance ICs [9]-[10]. Moreover, as compared to other leakage components namely gate leakage and BTBT leakage [11]-[12], it is most sensitive to parameter variations [12] because of its exponential dependence on effective channel length and temperature [12] - [14]. The analytical models presented under this framework are compared to BSIM models and are found to be more accurate as compared to existing statistical models [7]-[8]. We show in this paper that for accurate estimation of full-chip subthreshold leakage, it is important to consider die-to-die temperature variations which can significantly increase the leakage power due to electrothermal couplings involved between power and temperature [14]. Thus, temperature and power have been self-consistently evaluated to

ISLPED'04, August 9–11, 2004, Newport Beach, California, USA.

account for these couplings. Furthermore, the full-chip leakage power distribution arising due to both within-die and die-to-die P-T-V variations is calculated, which is subsequently used to estimate the leakage constrained yield under the impact of these variations.

The paper is organized as follows: In section 2, we present a mathematical framework to analyze the impact of single and multi-parameter variations. This framework is used to study the impact of within-die and die-to-die variations on subthreshold leakage in Sections 3 and 4. In Section 5, the leakage-constrained yield is estimated under the impact of variations. Concluding remarks are made in Section 6.

2. MATHEMATICAL PROBLEM FORMULATION

The electrical performance of a circuit is a function of environmental factors such as supply voltage (V_{dd}), temperature (T) etc., and physical parameters such as channel length (L), oxide thickness (t_{ox}) etc. This dependence of electrical performance (y) can be represented by $y=g(x_1, x_2 \ldots x_n)$, where $x_1, x_2 \ldots x_n$ are random variables representing the variations in above parameters. The nominal value (η_y), mean value (μ_y) and variance (σ_y) of y can be expressed as (1)-(3):

$$n_y = g(\mu_{x_1}, \mu_{x_2} \ldots \mu_{x_n}) \tag{1}$$

$$\mu_y = \int_{x_n^{min}}^{x_n^{max}} \ldots \int_{x_1^{min}}^{x_1^{max}} g(x_1, x_2 \ldots x_n) p(x_1, x_2 \ldots x_n) dx_1 \ldots dx_n \tag{2}$$

$$\sigma_y^2 = \int_{x_n^{min}}^{x_n^{max}} \ldots \int_{x_1^{min}}^{x_1^{max}} \left(g(x_1, x_2 \ldots x_n) - \mu_y\right)^2 p(x_1, x_2 \ldots x_n) dx_1 \ldots dx_n \tag{3}$$

In the above equations, μ_{x1} , μ_{x2} ... μ_{xn} represent the mean and $p(x_1, x_2 \ldots x_n)$ represents the joint probability density function (*p.d.f.*) of random variables $x_1, x_2 \ldots x_n$. In addition to the above parameters, we are interested in *delta (δ)* and *spread (s)* defined as:

$$\delta_y = \frac{\mu_y - n_y}{n_y} \qquad and \qquad S_y^2 = \frac{\sigma_y^2}{n_y^2} \tag{4}$$

The parameter delta indicates a normalized increase while the spread indicates a normalized variance of y. The analysis presented in this paper assumes Gaussian distributions for $x_1, x_2 \ldots x_n$ which is supported by experimental observations [15]. In the next subsection, a mathematical framework to model the fluctuations in y under Gaussian variations in a single parameter is first presented. Next, a few functions ($y = g(x)$) relevant for our work have been analyzed under this framework (Table 1.) Subsequently, a framework to extend this methodology for independent multivariate Gaussian parameter variations is presented.

2.1. Single Gaussian Parameter Variation

Given a function $y=g(x)$, the mean of y can be expressed as:

$$\mu_y = \int_{-\infty}^{+\infty} g(x) f_{\mu,\sigma}(x) dx \tag{5}$$

where $f_{\mu,\sigma}(x)$, μ and σ represent the *p.d.f.*, mean and variance of x. Applying Taylor's theorem, we get:

$$\mu_y = \sum_{n=0}^{+\infty} \frac{g^{(n)}(\mu)}{n!} \int_{-\infty}^{+\infty} (x-\mu)^n f_{\mu,\sigma}(x) dx \tag{6}$$

Equation (6) can be simplified as:

$$\mu_y = \sum_{k=0}^{+\infty} \frac{g^{(2k)}(\mu)}{2^k k!} \sigma^{2k} = n_y + \sum_{k=1}^{+\infty} \frac{g^{(kn)}(\mu)}{2^k k!} \sigma^{2k} \tag{7}$$

By neglecting higher order terms in (7), we get:

$$\delta_y = \frac{\mu_y - n_y}{n_y} = \frac{g''(\mu) g(\mu)}{2} s^2 \tag{8}$$

From the above equation, it can be observed that the sign of δ depends on the second derivative of $g(x)$. A positive δ indicates an average increase in the parameter y. Table 1 below lists *delta* for different functions $g(x)$ that will be used in the later sections.

Table 1: δ_y for y = g(x) assuming Gaussian distribution for random variable x (with spread S).

y= g(x)	Delta (δ_y)
$y = ae^{\beta \frac{x}{\mu}}$	$\delta_y = e^{\frac{\beta^2}{2} S^2} - 1$
$y = \frac{A}{x} e^{-\beta \frac{x}{\mu}}$	$\delta_y \approx \left(1 + \beta S^2\right) e^{\frac{\beta^2}{2} S^2} - 1$, see note [1]

2.2. Multiple Gaussian Parameter Variation

In a manufacturing process, process variations can be expressed in terms of variations in parameters such as L, t_{ox}, V_{dd}, T etc. If these parameters are determined at different steps of the manufacturing process, they can be assumed to be statistically independent [15]. In this case, we assume that these Gaussian variables are perfectly uncorrelated. If the function $y=g(x_1, x_2 \ldots x_n)$ can be expressed in a variable separable form, expression for δ can be easily evaluated. For instance, if y can be expressed as a product of individual functions $g(x_i)$ i.e. if

$$y = \prod_{i=1}^{n} g_i(x_i) \quad then \quad \delta_y = \left[\prod_{i=1}^{n} \left(1 + \delta_{g_i}\right)\right] - 1 \tag{9}$$

Equation (9) is used in Section 3 for calculating δ_y under multi-variable parameter variations.

3. ANALYSIS OF SUBTHRESHOLD LEAKAGE UNDER WITHIN-DIE PARAMETER VARIATIONS

3.1. Calculation of Delta for I_{ds}

The subthreshold leakage current for a MOSFET can be modeled as [13]:

$$I_{ds} = \mu_{eff} C_{ox} \frac{W_{eff}}{L_{eff}} (m-1) V_T^2 e^{\frac{V_{gs} - V_{th}}{m V_T}} \left(1 - e^{-Vds/V_T}\right)$$

$$\approx I_{s0} \frac{W_{eff}}{L_{eff}} e^{\frac{V_{gs} - V_{th}}{m V_T}} \quad \text{where} \quad m = 1 + \frac{\sqrt{\varepsilon_{si} q N_a / 4\psi_B}}{C_{ox}} \tag{10}$$

Here, μ_{eff} is the effective mobility, C_{ox} is the gate-oxide capacitance, L_{eff} is the effective channel length, W_{eff} is the effective width, V_T is the thermal voltage, N_a is the channel doping concentration and ψ_B is the difference between the Fermi potential and the intrinsic potential. Since $\mu_{eff} \propto T^{-1.5}$ and $V_T \propto T$, I_{s0} in equation (10) can essentially be assumed temperature independent [13]. In other words, it can be regarded as a technology parameter, independent of device parameters such as effective channel length, supply voltage, and temperature. Now, for small variations in channel length, threshold

[1] Both functions are only valid within 3σ range around the mean value. The result is an approximation obtained by applying Talyor expansion on the *1/x* term, preserving only first two terms, and integrating x over 3σ range of the random variation.

voltage (V_{th}) can be assumed to be linear around nominal value of channel length (η_{Leff}) and can be expressed as:

$$V_{th} = \overline{V_{th}} - \beta_{L_{eff}} \frac{\Delta L_{eff}}{\eta_{L_{eff}}} \tag{11}$$

where $\overline{V_{th}}$ is the nominal value of threshold voltage, $\Delta L_{eff}=\eta_{Leff}-L_{eff}$ and β_{Leff} is a constant for a device (NMOS/PMOS) defined as:

$$\boxed{\beta_{L_{eff}} = \frac{\eta_{L_{eff}}}{mV_T} \frac{dV_{th}}{dL_{eff}}} \tag{12}$$

Using (10) and (11), we can write,

$$I_{ds} = \frac{A}{L_{eff}} e^{-\beta_{L_{eff}} \frac{L_{eff}}{\eta_{L_{eff}}}}, \text{ where } A = I_{s0} W_{eff} e^{\frac{V_{gs}-\overline{V_{th}}}{mV_T}+\beta_{L_{eff}}} \tag{13}$$

We can now apply the result from Table 1 and calculate *delta* of leakage due to L_{eff} variation as:

$$\boxed{\delta_{I_{DS}} \approx \left(1+\beta_{L_{eff}} \cdot S_{L_{eff}}^2\right) e^{\frac{\beta_{L_{eff}}^2}{2} S_{L_{eff}}^2} - 1} \tag{14}$$

where S_{Leff} is the spread in channel length. Thus using (14), we can easily calculate the increase in subthreshold leakage current due to within-die channel length variations. The only unknown in (14) β_{Leff}, which is a critical parameter, will be estimated in the next section. The model presented above for calculating subthreshold leakage increase is simple but still accurate since BSIM equations are used to derive β_{Leff}.

3.2. Calculation of β

Analytically, β_{Leff} can be calculated using (12). The calculation of β for other variables such as T, t_{ox} and V_{dd} can be done using equations similar to (12). The threshold voltage (V_{th}) in (12) can be expressed as [10], [13], [16]:

$$V_{th} = V_{th0} + \Delta V_{th}(BODY) + \Delta V_{th}(HALO) - \Delta V_{th}(SCE, DIBL) \tag{15}$$

The fringe field effects and narrow-width effects have been ignored in (15). Vertical non-uniform doping (retrograde doping) is used to increase threshold voltage and it is considered a part of V_{th0}. $\Delta V_{th}(BODY)$ and $\Delta V_{th}(HALO)$ are independent of effective channel length whereas $\Delta V_{th}(SCE, DIBL)$ is exponentially dependent on channel length [13]:

$$\Delta V_{th}(SCE, DIBL) = \left[2(V_{bi}-\Phi_s)+V_{DS}\right] \cdot \left(e^{\frac{L_{eff}}{2L_t}} + 2e^{\frac{L_{eff}}{2L_t}}\right) \tag{16}$$

From (15) and (16), it can be observed that as effective channel length scales down, β defined by (12) grows exponentially thus significantly increasing the leakage. Because of an exponential dependence of $\delta_{I_{DS}}$ on β, the value of β is critical and hence BSIM equations are used to get an accurate estimate. We first use BPTM parameters [17] and BSIM3.2 equations [16] to calculate V_{th} as a function of L_{eff}, and (12) is then used to evaluate β. It was found β varies over a small range as shown in Figure 1. Hence, BSIM equations are used to calculate $\delta_{I_{DS}}$ as a function of effective channel length spread which is then curve-fitted with (14) to calculate an average value of β. Curve fitting allows us to consider the fact that β is slightly higher when effective channel length is shorter than normal, and transistors with shorter effective channel length contribute more to total leakage than transistors with longer effective channel length. In the above analysis, we have used BPTM [17] parameters for 100 *nm* (V_{DS}=*1.2V,* T_{nom}=*300K,* L_{eff}=*60nm*). It is shown in Section 3.3 that calculating β by above methodology and inserting its value in (14) gives results similar to those predicted by BSIM models.

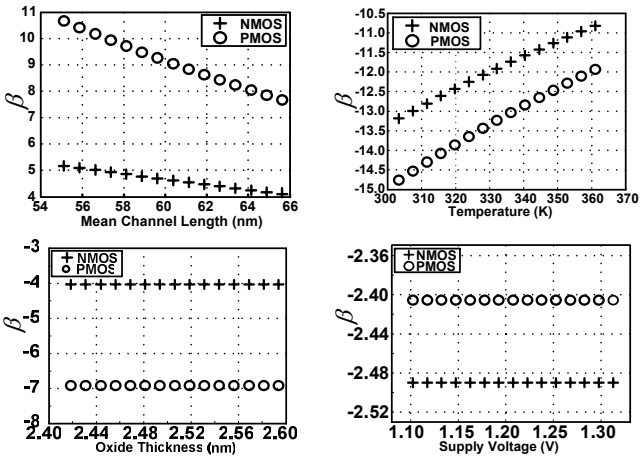

Figure 1: Variation of β with L_{eff}, T, t_{ox} and V_{dd}.

3.3. Comparison with Existing Statistical Models

In this section, we compare the results obtained under the proposed framework with those obtained from existing statistical models and BSIM model. The methodologies described in [7]-[8] are compared based on (13). Equation (17) is used to calculate the subthreshold leakage increase based on BSIM model, which assumes Gaussian variation for channel length

$$\delta_{I_{DS}} = \frac{\frac{1}{a} \cdot \int_{\mu-3\sigma}^{\mu+3\sigma} I_{SUB}(L) \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(L-\mu)^2}{2\sigma^2}} dL}{I_{SUB}(\mu)} - 1 \tag{17a}$$

where

$$a = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mu-3\sigma}^{\mu+3\sigma} e^{\frac{-(L-\mu)^2}{2\sigma^2}} dL \tag{17b}$$

In the above equation, $I_{SUB}(L)$ is the subthreshold leakage current estimated using BSIM equations [16], $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ being the mean and the variance of the channel length. In this work, $I_{SUB}(L)$ is calculated by porting BSIM3.2 equation into MATLAB so that V_{th}, I_{ds} and other parameters can be easily calculated based on BSIM parameters.

Table 2: Comparison with existing statistical models

Model	$\delta_{I_{DS}}$
Our Model	$\delta_{I_{DS}} \approx \left(1+\beta_{L_{eff}} \cdot S_{L_{eff}}^2\right) e^{\frac{\beta_{L_{eff}}^2}{2} S_{L_{eff}}^2} - 1$
Narendra et al. [7]	$\delta_{I_{DS}} = e^{\frac{\beta_{L_{eff}}^2}{2} S^2} - 1$
Srivastava et al. [8]	$\delta_{I_{DS}} = 2S^2 + 2\beta_{L_{eff}} S^2 + \beta_{L_{eff}}^2 S^2$

Table 2 above summarizes the increase in subthreshold leakage predicted by our model and the existing models [7]-[8]. Figure 2 plots the percentage increase in subthreshold leakage as a function of spread in channel length. Also, the increase in leakage estimated by BSIM models is plotted. The expression for [7] was derived by analyzing their I_{ds} model under our methodology, while the expression for [8] was derived by using the mean values for I_{ds} presented in [8]. It can be seen that our model compares well with BSIM model in the region of interest. On the other hand, Narendra et al. [7] underestimates the increase in leakage since it neglects the $1/L_{eff}$ pre-factor in (13). Srivastava et al. [8] overestimates the leakage

since it uses a Taylor series approximation for calculation of mean. It should be noted that since curve fitted values of β are used while comparing [7], results obtained under [7] will appear even worse if nominal values of β are used. This is because, the curve-fitting of β allows us to consider the variations in β which have been ignored in [7]-[8]. The model presented in [6] involves a large number of parameters and unlike (14), the expressions derived in [6] do not give a clear insight on the impact of variations on leakage. For instance, (14) clearly shows that variations result in an increase of leakage which cannot be easily discerned from [6].

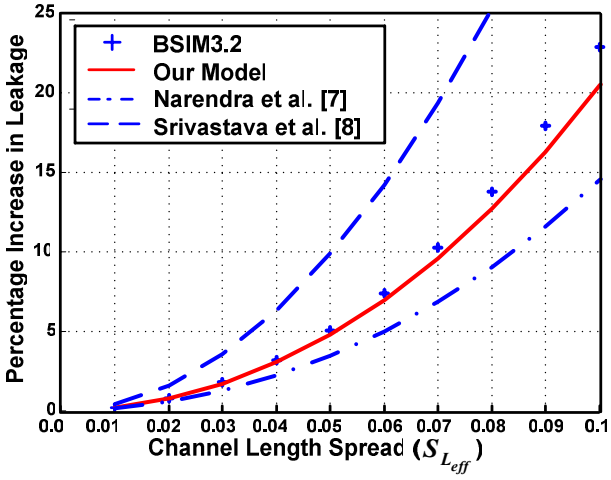

Figure 2: Percentage increase in leakage plotted for different values of $S_{L_{eff}}$ (for NMOS) predicted by different models.

The subthreshold leakage for a PMOS has stronger dependence on channel length variation as compared to an NMOS as shown in Figure 3. This is because of the fact that NMOS has β of 5.2 as compared to a β of 10 for PMOS at 100 nm technology node. This is due to a steeper V_{th} roll off slope for PMOS than for NMOS [16].

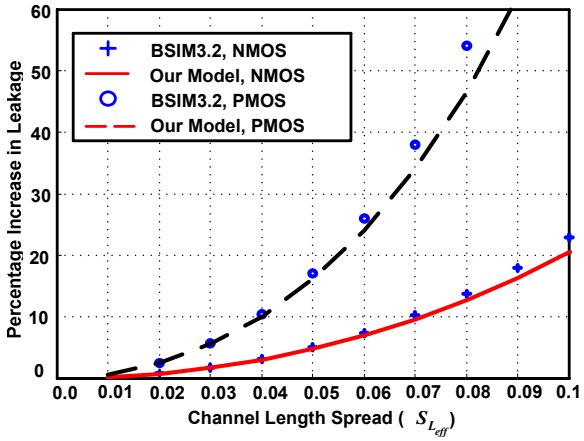

Figure 3: Percentage increase in leakage plotted for different values of $S_{L_{eff}}$ for NMOS and PMOS at 300K.

3.4. Considering all Within-Die Variations

In addition to within-die channel length variation, we take into account within-die gate oxide thickness (t_{ox}), supply voltage (V_{dd}) and temperature (T) variations and assume them to be independent. Considering these variations, we can rewrite (10) as:

$$I_{ds} = I_{s0}\frac{W_{eff}}{L_{eff}} e^{\frac{V_{gs}-\overline{V_{th}}}{mV_T} - \beta_{L_{eff}}\frac{\Delta L_{eff}}{\overline{L_{eff}}} - \sum_{X:r.v.}\beta_X\frac{\Delta X}{\overline{X}}} \tag{18}$$
$$= I_{s0}W_{eff}e^{\frac{V_{gs}-\overline{V_{th}}}{mV_T}}\left(\frac{1}{L_{eff}}e^{-\beta_{L_{eff}}\frac{\Delta L_{eff}}{\overline{L_{eff}}}}\right)\prod_{X:r.v.}\left(e^{-\beta_X\frac{\Delta X}{\overline{X}}}\right)$$

Here, X represents the random variable for different parameters such as T_{ox}, T, and V_{dd}, $\overline{X}$ represents its nominal value, $\Delta X = \eta_X - X$ and β_X is a constant defined in (19). Here again, threshold voltage (V_{th}) is assumed to vary linearly around the nominal value of X.

$$\beta_X = \frac{\eta_X}{mV_T}\frac{dV_{th}}{dX} \tag{19}$$

Using (9), we can calculate the *delta* of the leakage current as:

$$\delta_{I_{ds}} \approx \left[\left(1+\beta_{L_{eff}}\cdot S^2_{L_{eff}}\right)e^{\frac{\beta^2_{L_{eff}}}{2}S^2_{L_{eff}}}\right]\prod_{X:r.v.}\left(e^{\frac{\beta^2_X}{2}S^2_X}\right)-1 \tag{20}$$
$$\approx\left(1+\beta_{L_{eff}}\cdot S^2_{L_{eff}}\right)e^{\frac{\beta^2_{L_{eff}}}{2}S^2_{L_{eff}}}-1+\sum_{X:r.v}\left(e^{\frac{\beta^2_X}{2}S^2_X}-1\right)$$

According to ITRS 2003 [10], the effective channel length spread and gate oxide thickness spread are projected to stay at *3σ=10%* and *3%* respectively. From [4], we get an estimate of within-die supply voltage and within-die temperature variations, and choose supply voltage spread to be *3σ=5%* and temperature spread to be *3σ=3%* (*3σ* corresponds to 12K variation at a nominal value of 400K). Assuming above variations, Figures 4 and 5 plot β and δ for NMOS as well as PMOS. β_X is evaluated by a similar approach as in Section 3.2.

Note that threshold voltage is dependent on supply voltage variations through body effect. It can be observed that the impact of channel length and temperature variations is much greater than other variations. However, it is apparent from (20) that all of those variations serve to increase the total subthreshold leakage. In this analysis, we assumed that within-die temperature variations are uncorrelated with other within-die process variations, which is reasonable as long as within-die temperature gradient is determined from layout of different components in the die [4].

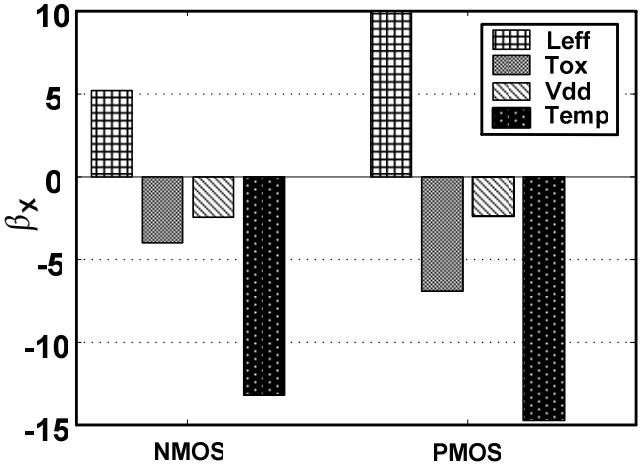

Figure 4: β for different within-die variations for NMOS and PMOS for 100 nm device at 300K.

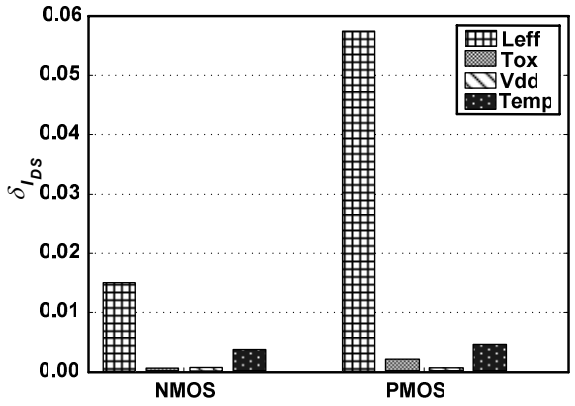

Figure 5: δ for leakage contributed by different within-die variations for NMOS and PMOS at 300K.

4. Analysis of Subthreshold Leakage under Die-to-Die Parameter Variations

Die-to-die variations include die-to-die channel length, temperature and voltage variations. For the purpose of yield estimation, ICs are generally screened assuming a worst-case supply voltage. Moreover, now-a-days, state-of-the-art voltage regulators are relatively insensitive to the drawn current. Therefore, we can neglect die-to-die voltage variations in our analysis. On the other hand, die-to-die temperature variations are a function of total chip power and correlate strongly to with-in-die process variations such as within-die channel length variations. Therefore, we will calculate the total chip power and die-temperature self-consistently as in [14] to study the impact of die-to-die temperature variations. Die-to-die channel length variations will be taken into account by varying the mean value of channel length. To illustrate how die-to-die and within-die variation impact subthreshold leakage distribution, we calculate leakage distribution using both BSIM3.2 simulations and analytical calculations for three cases. In case 1, we consider only the most significant die-to-die process variation i.e., effective channel length variation (3σ=5%). Equations (21a) and (21b) are used for calculating the leakage for BSIM and analytical calculations.

For $\mu_{L_{eff}} - 3\sigma_{die-to-die,L_{eff}} \le L_{eff} \le \mu_{L_{eff}} + 3\sigma_{die-to-die,L_{eff}}$,

$$I_{1Leak,BSIM}(L_{eff}) = W_n.I_{SUB,n}(L_{eff}) + W_p.I_{SUB,p}(L_{eff}) \quad (21a)$$

$$I_{1Leak,Ana}(L_{eff}) = I_{1Leak,Ana,n}(L_{eff}) + I_{1Leak,Ana,p}(L_{eff})$$
$$= W_n.\frac{A_n}{L_{eff}} e^{-\beta_{L_{eff},n}\frac{L_{eff}}{\eta_{L_{eff}}}} + W_p.\frac{A_p}{L_{eff}} e^{-\beta_{L_{eff},p}\frac{L_{eff}}{\eta_{L_{eff}}}} \quad (21b)$$

In the above equations $I_{SUB,n}$ and $I_{SUB,p}$ are subthreshold currents calculated through BSIM. Equation (21b) is based on (13). In case 2, apart from die-to-die channel variations, we consider within-die variations as well. Among within-die variations, we consider channel length and temperature variations only since they are most significant (Figure 5.) We consider 10% within-die variations on top of die-to-die variations. Equations (22a) and (22b) are used for BSIM and analytical calculations.

$$I_{2Leak,BSIM}(L_{eff}) = W_n.I_n(L_{eff}) + W_p.I_p(L_{eff}) \quad (22a)$$

$$I_{2Leak,Ana,n} = I_{1Leak,Ana,n}\left(1 + \beta_{L_{eff},n} \cdot S^2_{within-die,L_{eff}}\right) \times \left(e^{\frac{\beta_{L_{eff},n}^2}{2} S^2_{within-die,L_{eff}}} . e^{\frac{\beta_T^2}{2} S^2_{within-die,T}}\right)$$

$$I_{2Leak,Ana}(L_{eff}) = I_{2Leak,Ana,n}(L_{eff}) + I_{2Leak,Ana,p}(L_{eff}) \quad (22b)$$

In the above equations, I_n and I_p are subthreshold currents calculated using BSIM assuming Gaussian distribution for channel length and temperature. $I_{2Leak,Ana,p}$, which represents analytically estimated subthreshold current, is calculated in a similar way as $I_{2Leak,Ana,n}$. In case 3, we consider die-to-die temperature variations apart from other variations considered in case 2. Die-to-die temperature variations arise because of the couplings involved between power dissipation and die temperature. These couplings are taken into account by self-consistently evaluating the temperature as in [14]:

$$P_{Total}(L_{eff}) = I_{2Leak}(L_{eff}).V_{DD} + P_{active}$$
$$T \Rightarrow T + \theta.P_{Total} \quad (23)$$

Here I_{2Leak} is calculated from (22a) or (22b). In the calculation, we assume a microprocessor consisting of 96 million gates and average total W/L per NMOS-PMOS pair was taken to be 15. We use BPTM specified parameters in our analysis [17]. The results for the above three cases at 300K are shown in Figure 6. The X-axis plots the channel length variation (due to die-to-die channel length variations) and Y axis plots the leakage power.

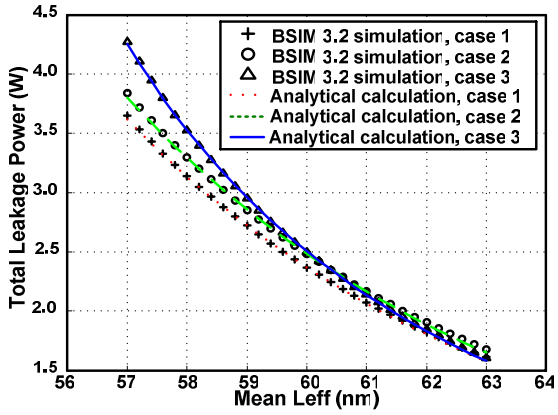

Figure 6: Total subthreshold leakage power vs. mean die-channel length at die temperature of 300K.

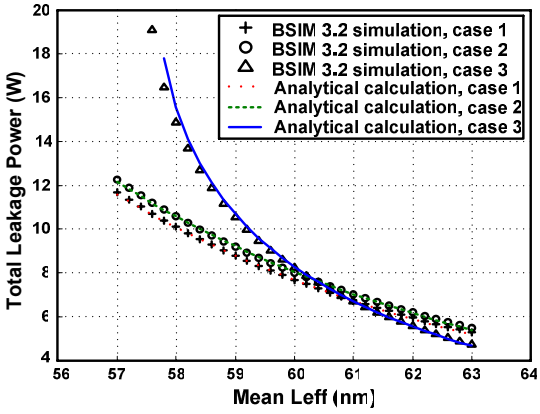

Figure 7: Total subthreshold leakage power vs. mean die-channel length at die temperature of 320K

Figure 7 compares the three cases at a higher die temperature of 320K (47˚C). It can be observed that at an elevated die temperature, leakage could be significantly larger than at 300K. Also, it can be seen that within-die variations increase the leakage without increasing the spread unlike die-to-die temperature variations. Further-

more, chips with low average channel length will have high leakage power, high temperature, and subsequently, even higher leakage power because of temperature-power couplings.

5. YIELD ESTIMATION

We now estimate the distribution of total power across different dies which can be used further to calculate the yield. The Gaussian probability density function $p(P_{Total})$ can be calculated using:

$$p(P_{total}) = -p(L_{eff})\left(\frac{dP(L_{eff})}{dL_{eff}}\right)^{-1}\Bigg|_{P(L_{eff})=P_{total}} \quad (24)$$

Here, $p(L_{eff})$ represents the *p.d.f.* for L_{eff} and $P(L_{eff})$ denotes the total power at a mean effective channel length of L_{eff}.

Figure 8 plots the probability density function for total leakage power based on (24). As we take more variations into account, the spread of the total leakage power distribution increases which implies that larger number of dies have higher leakage power. Since active power is relatively insensitive to variations and can be assumed constant, the yield can be defined by the number of dies having the leakage to active power ratio less than a maximum allowable ratio (say *r*). Figure 9 plots the yield as a function of *r* for three different cases.

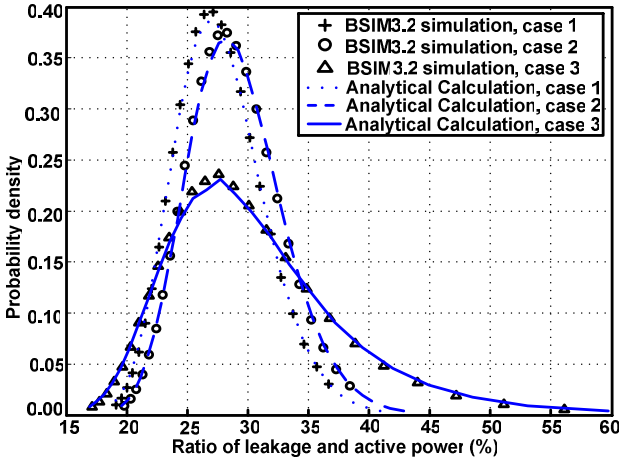

Figure 8: Leakage power probability density vs. ratio of leakage power and active power at 320K.

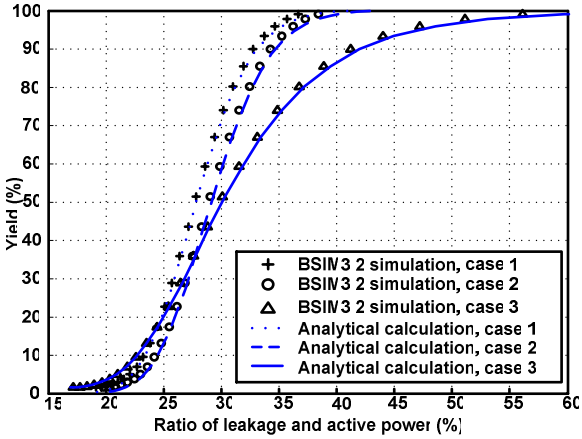

Figure 9: Yield vs. ratio of leakage power and active power at 320K.

It can be clearly observed that variations always result in a lower yield. In particular, die-to-die temperature variations due to electrothermal couplings between power and temperature significantly lower the yield. For instance, if r is 35%, then yield is reduced from 95% in case 1 to 91% in case 2 and to 72% in case 3. Therefore, to get an accurate estimate of yield, within-die variations as well as die-to-die variations should be taken into account.

6. CONCLUSION

A novel probabilistic approach to analyze the impact of within-die and die-to-die process (P), temperature (T) and voltage variations (V) on subthreshold leakage has been presented which allows an understanding of the relative sensitivities of subthreshold leakage to P-T-V variations. The results obtained are compared with BSIM results and are found to be more accurate as compared to those obtained from existing statistical models. Also, it has been shown that die-to-die temperature variations significantly increase the leakage due to the electrothermal couplings between subthreshold leakage power and temperature, especially at higher operating temperatures. Furthermore, the leakage power distribution arising due to within-die and die-to-die P-T-V variations has been calculated, which is subsequently used to estimate the leakage constrained yield under the impact of these variations. Also, it was shown that ignoring within-die and die-to-die process and temperature variations can lead to significant errors in yield estimation, which is further expected to degrade with technology scaling.

ACKNOWLEDGEMENTS

This work was supported by Intel Corp., Fujitsu Labs of America, and the University of California-MICRO program. The authors would also like to thank Dr. Vivek De and Dr. Frank O'Mahony of Intel's Circuit Research Lab for providing useful feedback.

REFERENCES

[1] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, V. De, "Parameter variations and impact on circuits and microarchitecture," *DAC*, 2003, pp. 338-342.
[2] S. Borkar, "Design challenges of technology scaling," *IEEE Micro*, vol. 19, pp. 23-29, Jul-Aug 1999.
[3] K. A. Bowman, S. G. Duvall, and J. D. Meindl, "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *IEEE J. Solid-State Circuits*, vol. 37, pp. 183-190, Feb. 2002.
[4] H. Su, F. Liu, A. Devgan, E. Acar, and S. Nassif, "Full chip leakage estimation considering power supply and temperature variations," *ISLPED*, 2003, pp. 78-83.
[5] R. Rao, A. Srivastava, D. Blaauw, and D. Sylvester, "Statistical analysis of subthreshold leakage current for VLSI Circuits," *IEEE Trans. on VLSI Systems*, vol. 12, no.2, pp. 131-139, 2004.
[6] S. Mukhopadhyay and K. Roy, "Modeling and estimation of total leakage current in nano-scaled CMOS devices considering the effect of parameter variation," *ISLPED*, 2003, pp. 172-175.
[7] S. Narendra, V. De, S. Borkar, D. Antoniadis and A. Chandrakasan, "Full-chip sub-threshold leakage power prediction model for sub-0.18µm CMOS," *ISLPED*, 2002, pp. 19-23.
[8] Srivastava, R. Bai, D. Blaauw and D. Sylvester, "Modeling and analysis of leakage power considering within-die process variations," *ISLPED*, 2002, pp. 64-67.
[9] Y. S. Lin, C. C. Wu, C. S. Chang, R. P. Yang, W. M. Chen, J. J. Liaw, and C. H. Diaz, "Leakage scaling in deep submicron CMOS for SoC," *IEEE Trans. Electron Devices*, vol. 49, pp. 1034-1041, Jun. 2002.
[10] Intl. Technology. Roadmap for Semiconductors (ITRS): 2003:http://public.itrs.net/
[11] K. Roy, S. Mukhopadhyay, H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," *Proc. of the IEEE*, vol. 91, no. 2, pp. 305-327, Feb, 2003.
[12] S. Mukhopadhyay, A. Raychowdhury, and K. Roy, "Accurate estimation of total leakage current in scaled CMOS logic circuits based on compact current modeling," *DAC*, 2003, pp. 169-174.
[13] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices,* Cambridge Univ. Press, 1998.
[14] K. Banerjee, S. C. Lin, A. Keshavarzi, S. Narendra, and V. De, "A self-consistent junction temperature estimation methodology for nanometer scale ICs with implications for performance and thermal management", *IEDM*, 2003, pp. 887-890.
[15] J. Zhang and M. Styblinski, *Yield and Variability Optimization of Integrated Circuits*, Kluwer Academic Publishers, Boston, 1995.
[16] BSIM3 manual: http://www-device.eecs.berkeley.edu/~bsim3/ftpv323/Mod_doc/BSIM3v323_manu.tar
[17] http://www-device.eecs.berkeley.edu/~ptm/mosfet.html