Statistical Analysis of Subthreshold Leakage Current for VLSI Circuits

Rajeev Rao, *Student Member, IEEE*, Ashish Srivastava, *Student Member, IEEE*, David Blaauw, *Member, IEEE*, and Dennis Sylvester, *Member, IEEE*

***Abstract*—We develop a method to estimate the variation of leakage current due to both intra-die and inter-die gate length process variability. We derive an analytical expression to estimate the probability density function (PDF) of the leakage current for stacked devices found in CMOS gates. These distributions of individual gate leakage currents are then combined to obtain the mean and variance of the leakage current for an entire circuit. We also present an approach to account for both the inter- and intra-die gate length variations to ensure that the circuit leakage PDF correctly models both types of variation. The proposed methods were implemented and tested on a number of benchmark circuits. Comparison to Monte Carlo simulation validates the accuracy of the proposed method and demonstrates the efficiency of the proposed analysis method. Comparison with traditional deterministic leakage current analysis demonstrates the need for statistical methods for leakage current analysis.**

***Index Terms*—Estimation, leakage currents, Monte Carlo, probability, process variability.**

I. Introduction

THE prominence of leakage currents in modern integrated circuits (ICs) has been spurred by the continued scaling of both supply voltage (V_{dd}) and threshold voltage (V_{th}). The exponential relationship between V_{th} and leakage current ($I_{\rm off}$) is central to this problem since V_{th} must be reduced to maintain good device switching speeds at low supply voltages. With the proliferation of portable applications that spend significant time in standby mode, large $I_{\rm off}$ values become a critical roadblock to improved battery lifetimes [1]. For example, static power is estimated to account for 15%–20% of the total power budget in high-performance ICs at the 130-nm technology node [2] and a number of methods for leakage reduction have been proposed for standby mode and during run time [3]–[13].

In addition to the rapid growth of $I_{\rm off}$ with each technology generation due to its exponential dependency on V_{th}, the potential also exists for large fluctuations of $I_{\rm off}$ from die to die or even gate to gate within a die. This is particularly true since controlling V_{th} is made more difficult in nanometer scale MOSFETs by drain-induced barrier lowering (DIBL) and discrete dopant effects [14]. While DIBL has been a problem since channel lengths first reached submicron dimensions, it is exacerbated in sub-100-nm devices by fundamental scaling limitations on oxide thickness ($T_{\rm ox}$). Reductions in $T_{\rm ox}$ have kept DIBL at reasonable levels since the gate could also be more strongly coupled to the channel in this way. For $T_{\rm ox}$ values below 1.5 nm, gate oxide leakage effects become significant and limit the scalability of $T_{\rm ox}$. Discrete dopant effects are important only in very narrow devices at very advanced technologies but lead to potentially large random fluctuations in channel doping levels and, therefore, V_{th}. In a projected 50 nm technology, the V_{th} 3σ uncertainty due to discrete dopant effects is expected to be comparable to the magnitude of the nominal V_{th} itself [15].

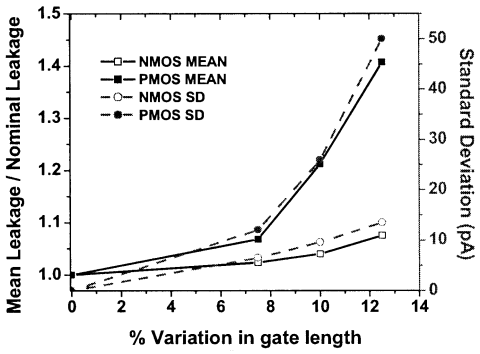


Fig. 1. Dependence of mean and standard deviation of leakage current on 3σ variation in gate length.

With the growing uncertainty in threshold voltage, estimation of $I_{\rm off}$ for a device becomes difficult, making the use of traditional delay-oriented corner models for leakage analysis impractical [16]. Worst case model files can easily exhibit 10–100$\times$ larger $I_{\rm off}$ than a nominal device, which leads to excessive guardbanding and overly conservative design practices. However, ignoring $I_{\rm off}$ variability altogether is also not an option: Consider a circuit block in which a small number of very leaky devices easily dominate the total static power consumption. Fig. 1 shows that the average leakage can be much larger ($\sim$30% for pMOS with L $3\sigma = 12.5\%$) than the nominal leakage due to the exponential dependence of current on the gate length. This observation also invalidates the use of nominal device model files for even typical dies. The results also 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 DIBL effects in pMOS devices are typically worse than in nMOS devices [17].

The above discussion points toward the statistical modeling of leakage current as a key unexplored area in future high-performance IC design. Monte Carlo simulations provide a method to analyze the effect of process variation, but are very expensive

Manuscript received March 1, 2003; revised June 30, 2003. This work was supported in part by the MARCO/DARPA Gigascale Silicon Research Center, National Science Foundation, and Semiconductor Research Corporation.

The authors are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA.

Digital Object Identifier 10.1109/TVLSI.2003.821549

in terms of time complexity. An analytical approach to leakage current estimation is therefore needed to enable the prediction of leakage power in a design before it has been fabricated [18].

A statistical leakage analysis method was previously proposed in [19] for modeling the impact of gate length variations, gate oxide thickness variations, and doping fluctuations on leakage. The analysis, however, was limited to stacks of single transistors and the extension to multi-transistor stacks is not straightforward. Also, the inter- and intra-die component of process variation was not accounted for. In this paper, we therefore propose a new analytical approach for statistical leakage estimation that can be applied to general circuit topologies and accounts for both inter- and intra-die variations.

As found in [19], the variation in gate length has the strongest impact among the process parameters affecting the leakage current. Gate oxide thickness is extremely well controlled by modern processes and the effect of the channel doping on the leakage current is fairly small. Hence, in this paper we only consider variation of the gate length. The dependence of the gate oxide thickness and channel doping can be expressed in the same form as the dependence on drawn gate length and, hence, the same approach can be adapted to include variability in these process parameters as well.

II. Analytical Approach to Leakage Variability

Our goal is to obtain an analytical model that allows efficient computation of the PDF of leakage currents for a circuit block or chip across the manufactured die. Using this model, we avoid the high computational costs of Monte-Carlo-based simulations that are impractical for analysis during the design process.

The objective is to find the PDF of the subthreshold leakage current I for a circuit block or chip, given the PDF of the drawn channel length L. We perform this task in three phases. First, we compute the leakage current distribution of individual gates in the circuit using the method described in Section II-A. Due to the exponential dependence of leakage current on gate length, the distribution of gate leakage for an individual gate has a lognormal shape. Secondly, based on the mean and variance of the leakage current distribution of individual gates, the total leakage of a circuit block is computed based on approximations for sums of lognormal distributions, as described in Section II-B. For this analysis, we assume that the channel lengths of all the gates are independent random variables. In the third phase, the impact of inter-die gate length variation is accounted for using a discrete PDF of the inter-die component of gate length variation. As described in Section II-C, the distribution of the leakage current due to intra-die variability is repeatedly computed, each time centered at a gate length that is shifted from the nominal value due to inter-die variation. We then take a weighted sum of this set of intra-die leakage distributions to obtain the total leakage distribution accounting for both inter- and intra-die gate length variations. Since the number of discretizations of the inter-die gate length PDF is typically small, the runtime of the third phase of the analysis remains small. We conclude Section II-C by justifying our earlier assumption of the independence of the channel lengths of the gates.

Although this paper deals with variability with respect to only the drawn channel length, the general framework can be easily used to model variability with respect to other process parameters. Furthermore, it is also feasible to use this approach to model the impact of process variability on other types of leakage current. For instance, to model the variability of gate tunneling current, we need to develop a new set of empirical functions to express leakage in terms of the oxide thickness $T_{\rm ox}$ since this is the process parameter that has the greatest influence on gate leakage. Thus, the proposed framework can be extended to model the overall impact of a number of different variation sources on the total leakage current of a chip.

A. Leakage Distribution of Individual Gates

We begin by describing our method for computing an analytical expression for the PDF of gate leakage of an individual gate.

First, the dependence of I on L is characterized by the function h such that $I = h(L)$. We then determine the inverse function $g(I)$, that expresses L as a function of I : $L = h^{-1}(I) = g(I)$. In order to compute the PDF of the leakage, it is essential that: 1) the function g is a closed-form expression and 2) the function h is differentiable over the given range of currents. Unfortunately, the complexity of the relationship between leakage current and channel length [i.e., the function $h(L)$] does not allow for the derivation of $g(I)$ such that it satisfies these two conditions. Therefore, as will be explained in Sections II-A and B, we propose an approximate fit for the function $h(L)$, such that the required inverse function can be computed while maintaining good accuracy.

Given the closed-form expression of $g(I)$ and the PDF of $L = f_x(L)$, we can express the PDF of I using the above expressions [20]

$$\text{PDF}(I) = f_y(I) = \frac{f_x\left(g(I)\right)}{h'(L)}. \tag{1}$$

Here, $h'(L)$ is the first derivative of the function $h(L)$. In our analysis, we assume that the drawn gate length has a Gaussian distribution with a fixed mean μ and standard deviation σ. Using these facts we can write the PDF of I as follows:

$$\begin{aligned}\text{PDF}(I) &= f_y(I) \\ &= \left(\frac{1}{h'(L)}\right)\left(\frac{1}{\sigma\sqrt{2\pi}}\right) \\ &\quad \times \exp\left(\frac{-\left(g(I)-\mu\right)^2}{2\sigma^2}\right).\end{aligned} \tag{2}$$

Finally, to calculate the mean and standard deviation of the leakage current distribution of the gate, we perform numerical integration of $f_y(I)$ over the given range of leakage currents

$$E(I) = \sum_I I.f_y(I) \tag{3}$$

$$SD(I) = \sqrt{\left(\sum_I I^2 f_y(I) - \left(\sum_I I\, f_y(I)\right)^2\right)}. \tag{4}$$

Below, we explain the proposed method for computing $f_y(I)$ in more detail for a single device. We initially discuss the approach for a single device and then extend the analysis for a stack of two or more transistors.

Single Transistor Stacks (Inverters): Based on the BSIM3v3 device model, the subthreshold current through a device can be expressed as [21]

$$I = I_0 \exp\left(\frac{(V_{gs} - V_{th})}{nV_T}\right)\left(1 - \exp\left(\frac{-V_{ds}}{V_T}\right)\right). \quad (5)$$

Here, $V_T = kT/q$ and $I_0 = \mu_0 C_{\text{ox}}(W_{\text{eff}}/L_{\text{eff}})V_T^2 \exp(1.8)$. The term $(1 - \exp(-V_{ds}/V_T))$ can be neglected for an inverter since $V_{ds} = V_{dd}$ is much greater than the thermal voltage V_T. We also set $V_{gs} = 0$ since the source nodes of either device in an inverter are tied directly to a supply rail and do not vary. V_{th} is the threshold voltage and is given by

$$V_{th} = V_{fb} + |2\phi_P| + \frac{\lambda_b}{C_{\text{ox}}}\sqrt{2qN\varepsilon_S(|2\phi_P + V_{sb}|)} - \lambda_d V_{ds} \quad (6)$$

where λ_b is the body-effect factor and λ_d is the DIBL coefficient. Expressions for λ_b and λ_d are given as in (7) and (8) [21], [22], shown at the bottom of the page. Here, V_{fb} is the flat-band voltage, V_{sb} is the source-to-body (substrate bias) voltage, N is the channel doping concentration, X_j is the junction depth, and W_{sd} is the sum of the depletion depths of the source and drain (reflecting the channel doping profile and biases). In the results reported in this paper, we use a single set of core BSIM3v3 SPICE parameters, with process variation only affecting the channel length of each individual device. Although the second-order parameters in the BSIM model can change as the channel length deviates from nominal, we observe that in a common industrial process, TSMC 0.18 μm, the same set of BSIM3v3 SPICE parameters are used for (drawn) channel lengths up to 0.5 μm. For channel lengths of 0.5 μm to 1.2 μm, a different model (and a correspondingly different set of parameters) is used. In our experiments, we focus on investigating cases where the maximum variation is 15% from the nominal 0.18 μm. Since these changes in L due to process variation are much smaller than the range of the binned model, we use the same binned fitting parameters for devices experiencing this degree of process variation.

These equations in principle enable us to calculate these parameters using the device model files for a given technology. However, analytical expressions for leakage current based on these parameters were found to fit very poorly for 0.18 μm technologies. In particular, nebulous definitions for the values for technology constants such as N_{sub} and X_j produce large errors in the analytical current expressions. Further, (7) and particularly (8) are inadequate in modeling λ_b and λ_d [18] and produce unrealistically small values for these parameters resulting in large errors in the values for leakage current.

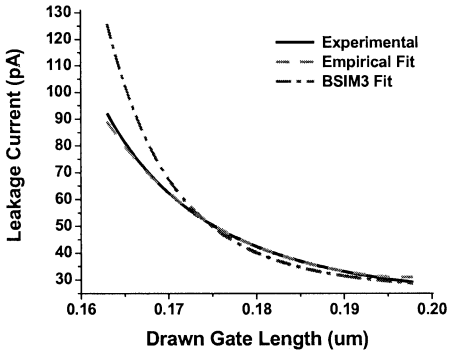


Fig. 2. Comparison of the BSIM3 fit and analytical fit for $h(L)$ with results from SPICE.

The actual BSIM3 model used to compute leakage current in SPICE simulations is much more complex than the simplified expressions presented in (5)–(8). Additionally, the constraints placed on functions g and h necessitates the use of further simplifications to derive a suitable analytical expression for current in terms of drawn gate length. Considering (6) in conjunction with (7) and (8) and keeping all parameters constant except L, we can rewrite V_{th} as a polynomial function of L_{eff}^{-1}. From Fig. 2 we see that this simplified BSIM3 model vastly overestimates the leakage current for devices with gate lengths that deviate by more than 5% from the nominal value. Since these conditions correspond to the devices that contribute a large portion of leakage current, the resulting PDF will be skewed to the right, rendering the BSIM3 fit unacceptable.

We therefore propose a new mathematical model to express leakage current I as a function of L.

$$I = q_1 \exp\left(q_2 L + q_3 L^2\right) = h(L). \quad (9)$$

This expression circumvents the use of V_{th} as an intermediate variable in expressing the current as a function of the gate length. However, it maintains the general form of the BSIM3 model since I is still expressed as a exponential function of a polynomial in L. The important properties of (9) are the following.

It preserves the exponential dependency of I on L.
It is easily invertible (as shown below).
It yields closed-form expressions for both I and L.
It accurately fits currents for both individual nMOS/pMOS as well as transistor stacks.

Fig. 2 shows the comparison between the values for leakage current obtained from SPICE simulations and the values obtained from both the BSIM3 fit and our empirical fit for a single

$$\lambda_b = 1 - \left(\sqrt{1 + \frac{2W_{sd}}{X_j}} - 1\right)\frac{X_j}{L_{\text{eff}}} \quad (7)$$

$$\lambda_d = \left[\frac{L_{\text{eff}}}{2.2\ \mu\text{m}^{-2}\,(T_{\text{ox}} + 0.012\ \mu\text{m})\,(W_{sd} + 0.15\ \mu\text{m})\,(X_j + 2.9\ \mu\text{m})}\right]^{-2.7} \quad (8)$$

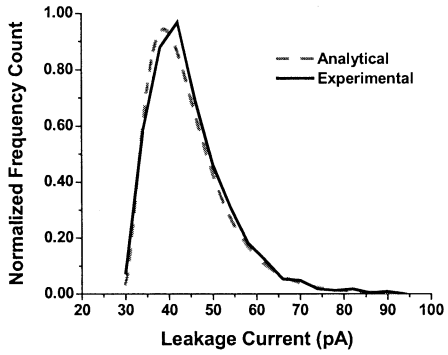


Fig. 3. Comparison of the SPICE PDF with the analytical PDF found from (2).

stacked device with $\pm 10\%$ variation in gate length. From the plot we can see that the empirical model provides a better fit over a wide range of channel lengths.

Equation (9) is a simple exponential quadratic equation that can be inverted to obtain an analytical expression for L as follows:

$$L = \left(\frac{1}{2q_3}\right)\left(-q_2 + \sqrt{q_2^2 - 4q_3 \ln\left(\frac{q_1}{I}\right)}\right) = g(I). \quad (10)$$

Using the expressions from (2)–(4) with the functions g, h as specified by (9) and (10), we can obtain the PDF of I. Fig. 3 presents the comparison between the PDF obtained from SPICE simulations and the PDF obtained analytically for a single stacked device with 10% 3σ variation in gate length. The plots of the PDF's, including the tail portion, match well and have a lognormal shape.

Series-Connected Devices (Stacks): In the case of a stack of transistors, the gate length variation impacts the leakage current of the bottom transistor in the stack in two ways: 1) gate length variation of the bottom transistor directly modulates its threshold voltage and 2) gate length variation of the top transistor indirectly affects the leakage of the bottom transistor by altering the voltage drop across the top transistors of the stack. Hence, the analytical expression of current as a function of gate length is more complex for stacks of multiple transistors. Since the devices in a stack are placed close together the layout, we make the simplifying assumption that their gate length variations in are perfectly correlated. Also, we derive the analysis for stacks of two and three transistors, the method can be extended to stacks of arbitrary length in a straightforward manner.In an inverter we ignored the term $(1 - \exp(-V_{ds}/V_T))$ in (5) since the drain–source voltage V_{ds} in the leaking device is much greater than the thermal voltage V_T. For a device with stacked structures of two or three transistors, the value of the intermediate node voltage (V_{ds2} and V_{ds3}) is much lower. In [23], the authors present a model to compute the drain-source voltage of transistors in stacks of arbitrary length. However, the complexity of these analytical expressions makes the derivation for a suitable equation for g [as in (1)] very difficult.

On the other hand, our empirical model is sufficiently robust to provide the leakage currents in these stacked circuits using the same general form of (9). The current is once again empirically modeled as

$$I = q_1' \exp\left(q_2' L + q_3' L^2\right) = h(L). \quad (11)$$

The constants q_1', q_2', q_3' are a new set of fitting parameters. Naturally, this set of constants will vary for different stack depths and also for nMOS versus pMOS since the drain–source voltages will differ. Equation (10) is then solved again using the suitable coefficients in the quadratic expression to obtain the value of channel length as a function of I and similarly the PDF can be determined.

A simplifying assumption generally made is treating ON transistors in a stack as short circuits [23]. This is a reasonable assumption when the ON device is not the top device in an nMOS stack or the bottom device in a pMOS stack. In these cases the ON devices lead to V_{th} drops from the nominal voltages, leading to an overall lower leakage current. We consider this effect by estimating the leakage current under the assumption that the V_{th} drop is a constant value that corresponds to the nominal V_{th} of the device. This allows us to use the same models for stacks of transistors with an effectively reduced power supply voltage.

B. Leakage Distribution of Circuit Blocks

In this section, we extend the approach developed to estimate the leakage current distribution for individual gates to the circuit level. Since the distribution of the leakage current of a single gate is close to lognormal, we approximate the leakage current for the circuit as a whole as the sum of lognormals. Thus, to find the distribution of the total leakage current, given k lognormal random variables (RV's) we need to find the distribution of the sum S given as

$$S = X_1 + X_2 + \cdots + X_k = e^{Y_1} + e^{Y_2} + \cdots + e^{Y_k}. \quad (12)$$

Sums of lognormals, assuming independence, can be well approximated by another lognormal RV [24]. Various approaches are known to estimate the parameters of the final lognormal. As shown in [24] a simpler Wilkinson approximation [25] is more accurate as compared to other complex approaches for our range of interest in the cumulative probability of leakage current. In Wilkinson's approach the sum of the mean and variance of the individual gate leakage current distributions, $X_1, X_2, \ldots, X_k$ is matched with the first two moments of S

$$E(S) = \mu_1 + \mu_2 + \mu_3 + \mu_4 + \cdots \quad (13)$$

$$Var(S) = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 + \cdots \quad (14)$$

where the μ's and σ's are the mean and standard deviation of the leakage currents of the individual gates. The PDF of a lognormal is given by

$$f(x) = \left(\frac{1}{x\sqrt{2\pi\beta}}\right) \exp\left(\frac{-(\ln(x) - \alpha)^2}{2\beta^2}\right) \quad (15)$$

where α and β are the parameters of the lognormal distribution. If Y (μ,σ) is a Gaussian random variable then the corresponding lognormal X is related to Y as $X = \exp(Y)$ and since the

parameters of the lognormal are the mean and variance of the corresponding Gaussian distribution, we need to compute these based on the mean and variance of the lognormal

$$E(X) = \exp\left(\alpha + \frac{\beta^2}{2}\right) \tag{16}$$

$$Var(X) = \exp[2(\alpha+\beta^2)] - \exp(2\alpha+\beta^2). \tag{17}$$

These values can be used to obtain the parameters of the Gaussian in terms of the mean and variance of the lognormal as given below

$$\alpha = \frac{1}{2}\log\left(\frac{E^4(X)}{E^2(X)+Var(X)}\right) \tag{18}$$

$$\beta^2 = \log\left(\frac{Var(X)+E^2(X)}{E^2(X)}\right). \tag{19}$$

The parameters of the lognormal are then obtained using (18) and (19), which completely determines the PDF of the leakage current of the circuit block. Note that for large circuit blocks the leakage current distribution will approach a Gaussian due to the Central Limit Theorem [20]. As shown in [26], both S (in (12)) as well as $\log S$ can be approximated by a Gaussian for large k. Thus, for large k, the lognormal distribution will tend toward the shape of a Gaussian distribution, and using a lognormal distribution to approximate sums of lognormals is justified.

C. Accounting for Inter- and Intra-Die Variations

Process variation can be classified into inter-die variation and intra-die variations. Intra-die variation refers to variation within a particular circuit block or chip. Inter-die variation occurs from one die to the next, meaning that the same device in the design has different features among different die. We consider the total drawn gate length of device i to be the algebraic sum of the nominal gate length $L_{nominal}$, the intra-die variation ΔL_{intra} and the inter-die variation ΔL_{inter}. Consequently, the total variance is a sum of the inter- and intra-die variances

$$L_{\text{total},i} = L_{nominal} + \Delta L_{\text{inter}} + \Delta L_{\text{intra},j} \tag{20}$$

$$\sigma^2_{\text{total}} = \sigma^2_{\text{inter}} + \sigma^2_{\text{intra}}. \tag{21}$$

In (20), the random variable ΔL_{inter} is shared by all devices in a design (creating correlation between their leakage currents), where as the random variables ΔL_{intra} assigned to each device are independent (reducing correlation of their leakage currents). ΔL_{inter} can also be interpreted as the distribution of the chip-mean gate length, where as ΔL_{intra} represents deviation in gate length of individual devices from this chip mean.

To compute the total leakage, accounting for both types of gate length variation, we first discretize the PDF of L_{inter} as shown in Fig. 4(a). For each discrete point $L_{\text{inter},j}$ on the PDF of L_{inter}, we consider the intra-die variation of the channel length as a normally distributed PDF, whose mean is $L_{\text{inter},j}$ and standard deviation is σ_{intra}. This distribution considers the variation in channel length due to intra-die variations with the mean being determined by the inter-die variability and the shape being determined by the intra-die variability. We then use the approach outlined in Sections II-A and B to obtain the PDF of the leakage

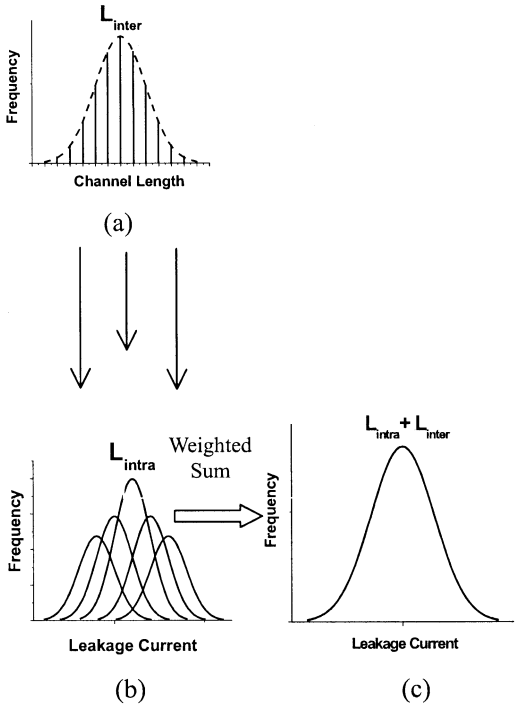


Fig. 4. PDFs for (a) channel length considering only L_{inter}, (b) leakage current corresponding to each point in (a) considering L_{intra}, and (c) leakage current considering both L_{inter} and L_{intra}.

current corresponding to each of the discrete points on the L_{inter} PDF as shown in Fig. 4(b). Thus, we obtain a family of the PDFs of leakage current, where each PDF is associated with a conditional probability that corresponds to the PDF value of $L_{\text{inter},j}$ on the PDF of L_{inter}. To obtain the PDF of leakage current considering both variations we form a weighted sum of the family of PDF's. This can be expressed as

$$P(I < i < I + \Delta I) = \sum_{j=1}^{n} P_{\text{intra},j} \times (I < i < I + \Delta I)^* P_{\text{inter}}(L_{\text{inter},j}) \tag{22}$$

where $P_{\text{inter}}(L_{\text{inter},j})$ is the probability of occurrence of jth point from the set of 'n' discrete points selected. P_{intra} is calculated based on the lognormal distribution of the leakage current corresponding to the jth point, $L_{\text{inter},j}$ on the L_{inter} PDF.

Intra-chip variations often exhibit spatial correlation such that devices that are closer to one another have a higher probability of being alike than devices that are far apart. In our analysis so far, we have assumed that the intra-die gate length variation expressed by the random variables ΔL_{intra} assigned to each gate is independent. However, spatial correlation will result in dependence of these random variables. Hence, we examine the impact of such correlation on the statistical leakage estimation using Monte Carlo simulation. For simplicity, we model the effect of spatial correlation using clusters of gates in a circuit, such that ΔL_{intra} of gates within a cluster are perfectly correlated, while ΔL_{intra} of gates between different clusters are independent. Large cluster sizes therefore reflect a stronger spatial correlation of intra-die gate length variation while small cluster sizes reflect a weak spatial correlation.

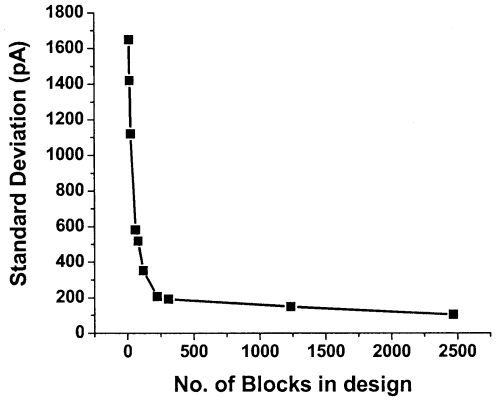


Fig. 5. Relation between the standard deviation of total leakage current in a chip and the number of blocks that constitute the chip sheds insight on the validity of gate independence assumptions.

In Fig. 5, we show the standard deviation of leakage current for a design as a function the number of clusters in the design. As the number of clusters is decreased, the size of each individual cluster increases, representing a stronger spatial correlation. From the plot, we see that due to the averaging effect of a large number of uncorrelated variables, the variability in leakage current converges to a relatively small value as the number of clusters is increased. For designs with 250 or more clusters, the standard deviation has largely converged, and the impact of spatial correlation can be ignored. In other words, comparing the case where all gates are considered to have independent intra-die gate length variation (as assumed in the analysis in this paper), to the case having 250 gate clusters with perfectly correlated intra-die gate length variation within each cluster, results in negligible error. In typical process technologies, spatial correlation drops off sharply for distances greater than 0.1 mm [27]. Hence, even for a small design with a die area of 2.5 mm^2, the number of independent gate clusters is sufficient to perform statistical leakage current analysis assuming independence of intra-die gate length variation. Since most practical designs are significantly larger than 2.5 mm^2, spatial correlation does not pose a significant issue for statistical leakage current estimation for such designs.

III. Results

In this section we first compare the results obtained from the analytical approach outlined in Section II and Monte Carlo simulations for individual gates and circuit blocks assuming only intra-die variation. We then present the comparison between SPICE simulations and our analytical approach for the ISCAS benchmark circuits considering both inter- and intra-die variation. We also show the difference between deterministic analysis and statistical analysis for various circuit blocks. In the analysis that follows we only compare the mean and variance of the two approaches.

Table I compares the analytical approach to Monte Carlo simulations for a single gate. The drawn gate length is assumed to be normally distributed, with the 3σ variation being 10% of the mean, the mean being 0.18 μm. The table shows that the error in estimating the mean leakage current varies from 0–6% but is typically $<4\%$. The estimate in standard deviation shows higher error for one of the cases, but in all other cases it is small. The leakage current can be seen to drop significantly while going from a 1-stack to a 2-stack which is due to the well-known "stack effect." The stack effect shows an even larger reduction for standard deviation when going from stacks of depth one to two. For the nMOS stacks, the mean reduces by a factor of 5 while the standard deviation reduces by a factor of 9.

Table II shows the results of the comparison of the analytical approach to Monte Carlo simulation for nine ISCAS85 benchmark circuits [28]. The 3σ variation in the drawn gate length is set at $\pm 10\%$. The experimental mean and standard deviation are calculated for a random set of input vectors for each circuit. The table shows that the average error in estimating the mean over all circuits is 3.5% with a maximum error of 8.3%. The average error in the standard deviation is 13.7% with a maximum error of 17.9%.

TABLE I
Comparison of the Analytical Approach With Monte Carlo Simulations For nMOS/pMOS Stacks

3σ Var = 10%							
		Mean (pA)			SD (pA)		
		Exp	Ana	%Error	Exp	Ana	%Error
1-Stack	PMOS	32.3	34.1	5.6	25.9	28.9	11.6
	NMOS	44.1	44.4	0.7	9.6	10.0	4.2
2-Stack	PMOS	4.7	4.7	0.0	0.3	0.3	0.0
	NMOS	9.1	9.1	0.4	1.0	1.0	3.1
3-Stack	PMOS	3.6	3.6	0.0	0.1	0.1	0.0
	NMOS	5.8	6.0	3.1	0.6	0.6	3.4

TABLE II
Comparison of the Analytical Approach With Monte Carlo Simulations for a Circuit Considering Only Intra-Die Variation

Circuit	Size (Gates)	Mean (nA)			SD (pA)		
		Exp	Ana	Err(%)	Exp	Ana	Err(%)
c17	6	0.2	0.3	8.3	36.0	37.0	2.8
c432	159	7.1	7.2	1.4	190.0	210.0	10.5
c499	519	19.0	20.0	5.3	280.0	330.0	17.9
c880	364	17.0	17.0	0.0	280.0	330.0	17.9
c1355	528	21.0	22.0	4.8	320.0	370.0	15.6
c1908	432	16.0	17.0	6.3	260.0	300.0	15.4
c2670	825	32.0	33.0	3.1	350.0	410.0	17.1
c3540	940	39.0	40.0	2.6	420.0	480.0	14.3
c6288	2470	120.0	120.0	0.0	900.0	1010.0	12.2

TABLE III
Comparison of the Analytical Approach With Monte Carlo Simulations for a Circuit Considering Both Intra- and Inter-Die Variation

Circuit	Mean (nA)			SD (nA)		
	Exp	Ana	Err(%)	Exp	Ana	Err(%)
c17	0.4	0.4	0.0	0.5	0.4	20.0
c432	10.0	10.0	0.0	9.2	7.6	17.4
c499	28.0	27.0	3.6	24.1	19.5	19.1
c880	24.6	23.9	2.8	21.2	17.4	17.9
c1355	32.2	30.6	5.0	30.2	23.9	20.9
c1908	23.6	23.3	1.3	21.9	17.5	20.1
c2670	48.2	45.4	5.8	41.3	33.7	18.4
c3540	57.5	54.5	5.2	47.4	38.2	19.4
c6288	186.7	175.4	6.1	183.5	152.0	17.2

TABLE IV
TRADITIONAL ANALYSIS ESTIMATES THE MEDIAN VALUES ACCURATELY BUT THE HIGHER PERCENTILES ARE VASTLY OVERESTIMATED

Circuit	Leakage Current (nA)					
	Traditional Analysis			Statistical Analysis (Analytical) / Ratio		
	50%	95%	99%	50%	95%	99%
c17	0.2	0.8	2.1	0.2 / 1	0.3 / 2.67	0.5 / 4.2
c432	6.7	21.2	49.8	6.7 / .96	11.6 / 1.83	26.2 / 1.90
c499	18.2	57.2	134.6	19 / .96	34.3 / 1.67	78.5 / 1.71
c880	16.1	50.2	116.8	16.8 / .96	34.1 / 1.47	77.4 / 1.51
c1355	20.0	67.0	162.7	20.7 / .97	51.2 / 1.31	113.2 / 1.43
c1908	15.5	48.3	112.5	16.2 / .96	37.8 / 1.28	80.4 / 1.40
c2670	30.7	96.9	227.3	31.3 / .98	72.3 / 1.34	160.8 / 1.41
c3540	37.3	112.9	258.9	38.4 / .97	.86 / 1.31	179.7 / 1.44
c6288	111.1	402.1	1010.0	115.2 / .96	306.2 / 1.31	710.4 / 1.42

In our experiments considering both inter and intra-die variation, we assume the total standard deviation to be equal to 15% of mean. Table III compares the results of the analytical approach to Monte Carlo simulation considering both intra- and inter-die variation. The table lists the data for the case where intra-die and inter-die process standard deviation have been assumed to be 10% and 11% of mean, respectively, which make up a total standard deviation of 15% variation based on (21). As can be seen, the error in the estimated mean is always within 6.1% and that for the standard deviation within 21%.

We compare the new analytical approach to a traditional deterministic approach, where all gates lengths are assumed to be perfectly correlated and hence have the same length. Table IV compares the median and the 95th/99th percentile points estimated using the traditional approach to the new statistical approach.

As can be seen, the traditional approach significantly overestimates the leakage for higher confidence points since all the devices are assumed to be operating at the pessimistic corner point. Since the relationship between the gate length and leakage current is monotonic, we find the median point as estimated by a traditional analysis to be very close to the nominal leakage current.

Fig. 6 shows the impact of varying the distribution of inter-die process variation on the PDF of the leakage current while keeping the standard deviation of the total gate length $\sigma_{\text{total}} = 15\%$ of the mean. The figure shows that when inter-die process variation is increased (and consequently the intra-die variation is decreased), the PDF tends to a lognormal shape. Note that for the case of no intra-die process variation, all gate lengths on a single die will be at their nominal values. Hence the PDF of this leakage current due to inter-die process variation alone should be similar to the PDF of the leakage current of a single gate which, as we know, can be closely approximated by a lognormal.

The figure suggests that, since leakage current is well characterized in terms of the I_{DDQ} values across die, the shape of this leakage current PDF can be a useful way to estimate the contribution of the inter-die or intra-die component to the total process variation.

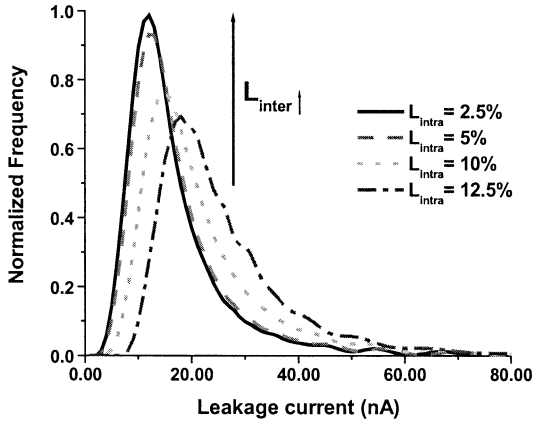


Fig. 6. PDFs of leakage current for different contributions of inter- and inter-die process variation. The total variation is 15%.

IV. CONCLUSION

In this paper, we have presented a method to estimate distributions of subthreshold leakage current in the presence of both inter- and intra-die process variations. We developed a model to predict leakage currents as a function of drawn gate length and have shown it to be fairly accurate over the range of values of interest. We then developed a new approach to estimate leakage currents PDFs in circuit blocks considering both inter- and intra-die process variation. We compared this approach to Monte Carlo simulations and have showed it to be accurate in estimating the overall mean and standard deviation of the leakage current in circuit blocks. We have shown that using the analytical approach we can significantly reduce the pessimism introduced by deterministic analysis while saving on the computational effort required for a Monte Carlo analysis. We have also highlighted the difference in the impact of inter and intra-die process variation on the PDF of leakage current.

ACKNOWLEDGMENT

The authors would like to acknowledge helpful discussions with S. Nassif from IBM ARL, Austin, TX.

REFERENCES

[1] C. Hu, "Device and technology impact on low power electronics ," in *Low Power Design Methodologies*, J. Rabaey, Ed. Norwell, MA: Kluwer, 1996, pp. 21–35.

[2] . [Online]. Available: http://developer.intel.com/design/mobile/datashts

[3] R. K. Krishnamurthy, A. Alvandpour, V. De, and S. Borkar, "High-performance and low-power challenges for sub-70 nm microprocessor circuits," in *Proc. CICC*, 2002, pp. 125–128.

[4] M. C. Johnson, D. Somasekhar, and K. Roy, "Models and algorithms for bounds on leakage in CMOS circuits," *IEEE Trans. Computer-Aided Design*, vol. 18, pp. 714–725, June 1999.

[5] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "1-V power supply high-speed digital circuit technology with multithreshold voltage CMOS," *IEEE J. Solid-State Circuits*, vol. 30, pp. 847–854, Aug. 1995.

[6] S. Sirichotiyakul, T. Edwards, C. Oh, J. Zuo, A. Dharchoudhury, R. Panda, and D. Blaauw, "Standby power minimization through simultaneous threshold voltage and circuit sizing," in *Proc. DAC*, 1999, pp. 436–441.

[7] Y. Ye, S. Borkar, and V. De, "A new technique for standby leakage reduction in high-performance circuits," in *Proc. Symp. VLSI Circuits*, 1998, pp. 40–41.

[8] J. Halter and F. Najm, "A gate-level leakage power reduction method for ultra-low-power CMOS circuits," in *Proc. CICC*, 1997, pp. 475–478.

[9] J. Kao, A. Chandrakasan, and D. Antoniadis, "Transistor sizing issues and tool for multi-threshold CMOS technology," in *Proc. DAC*, 1997, pp. 409–414.

[10] L. Wei, Z. Chen, K. Roy, Y. Ye, and V. De, "Mixed-V_{th} (MVT) cmos circuit design methodology for low power applications," in *Proc. DAC*, 1999, pp. 430–435.

[11] Q. Wang and S. Vrudhula, "Static power optimization of deep submicron CMOS circuits for dual V_T technology," in *Proc. ICCAD*, 1998, pp. 490–496.

[12] P. Antognetti, "CAD model for threshold and subthreshold conduction in MOSFET's," *IEEE J. Solid State Circuits*, vol. SC-17, pp. 454–458, June 1982.

[13] P. Pant, V. De, and A. Chatterjee, "Device-circuit optimization for minimal energy and power consumption in CMOS random logic networks," in *Proc. DAC*, 1997, pp. 403–408.

[14] "International Technology Roadmap for Semiconductors," Semiconductor Industry Association, 2001.

[15] D. J. Frank, R. Dennard, E. Nowak, P. Solomon, Y. Taur, and H. P. Wong, "Device scaling limits of Si MOSFET's and their application dependencies," *Proc. IEEE*, vol. 89, pp. 259–288, Mar. 2001.

[16] S. Narendra, V. De, S. Borkar, and A. Chandrakasan, "Full chip subthreshold leakage power prediction model for sub -0.18 μm CMOS," in *Proc. ISLPED*, 2002, pp. 139–23.

[17] S. Tyagi, M. Alavi, R. Bigwood, T. Bramblett, S. Tyagi, J. Bradenburg, W. Chen, B. Crew, M. Hussein, P. Jacob, C. Kenyon, C. Lo, B. McIntyre, Z. Ma, P. Moon, P. Nguyen, L. Rumaner, R. Schweinfurth, S. Sivakumar, M. Stettler, S. Thompson, B. Tufts, J. Xu, S. Yang, and M. Bohr, "A 130 nm generation logic technology featuring 70 nm transistors, dual V_t transistors and 6 layers of Cu interconnect," in *Proc. IEDM*, 2000, pp. 567–570.

[18] J. Kao, S. Narendra, and A. Chandrakasan, "Subthreshold leakage modeling and reduction techniques," in *Proc. ICCAD*, Nov. 2002, pp. 141–148.

[19] A. Srivastava, R. Bai, D. Blaauw, and D. Sylvester, "Modeling and analysis of leakage power considering within-die process variations," in *Proc. ISLPED*, 2002, pp. 64–67.

[20] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1991.

[21] A. Chandrakasan, W. J. Bowhill, and F. Fox, *Design of High-Performance Microprocessor Circuits*. Piscataway, NJ: IEEE Press, 2001.

[22] K. K. Ng, S. A. Eshraghi, and T. D. Stanik, "An improved generalized guide for MOSFET scaling," *IEEE Trans. Electron Devices*, vol. 40, pp. 1893–1895, 1993.

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

[24] N. C. Beaulieu, A. A. Abu-Dayya, and P. J. McLane, "Comparison of methods of computing lognormal sum distributions and outages for digital wireless applications," in *Proc. IEEE Int. Conf. Communications*, vol. 3, 1994, pp. 1270–1275.

[25] S. C. Schwartz and Y. S. Yeh, "On the distribution function and moments of power sums with lognormal components," *Bell Syst. Tech. J.*, vol. 61, pp. 1441–1462, Sept. 1982.

[26] N. Marlow, "A normal limit theorem for power sums of independent random variables," *Bell Syst. Tech. J.*, vol. 46, pp. 2081–2090, Nov. 1967.

[27] K Bernstein, personal communication, 2003.

[28] F. Brglez and H. Fujiwara, "A neutral netlist of 10 combinational benchmark circuits and a target translator in fortran," in *Proc. Int. Symp. Circuits and Systems*, May 1985, pp. 695–698.

Rajeev Rao (S'02) received the B.S. degree in electrical and computer engineering from Rutgers University, New Brunswick, NJ, in 2002. He is currently working toward the Ph.D. degree at the University of Michigan, Ann Arbor.

In summer 2003, he was with IBM Austin Research Lab, Austin, TX, working as a Research Co-op on leakage power analysis. His research interests include modeling and analysis of low-power VLSI designs and variability-aware circuit approaches.

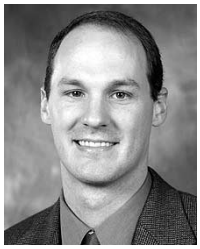
Ashish Srivastava (S'00) recieved the B.Tech. degree in electrical engineering in 2001 from the Indian Institute of Technology, Kanpur, India, and the M.S. degree in electrical engineering in 2003 from the University of Michigan, Ann Arbor, where he is currently working toward the Ph.D. degree.

In summer 2003, he was with the Technology CAD Division, Intel Corporation, Hillsboro, OR, where he was a Graduate Intern. His research interests include optimization and statistical analysis for high-performance VLSI design.

David Blaauw (M'93) received the B.S. degree in physics and computer science from Duke University, Durham, NC, in 1986, and the M.S. and Ph.D. degrees in computer science from the University of Illinois, Urbana, in 1988 and 1991, respectively.

He was a Development Staff Member at the Engineering Accelerator Technology Division, IBM Corporation, Endicott, NY, until August 1993. From 1993 to August 2001, he was with Motorola, Inc. Austin, TX, where he was the Manager of the High Performance Design Technology Group. Since August 2001, he has been an Associate Professor at the University of Michigan, Ann Arbor. His work has focused on VLSI design and CAD with particular emphasis on circuit analysis and optimization problems for high-performance and low-power designs.

Dr. Blaauw was the Technical Program Chair and General Chair for the International Symposium on Low Power Electronics and Design in 1999 and 2000, respectively, and was the Technical Program Co-Chair and Member of the Executive Committee for the ACM/IEEE Design Automation Conference in 2000 and 2001.



Dennis Sylvester (S'95–M'00) received the B.S. degree (*summa cum laude*) from the University of Michigan, Ann Arbor, in 1995, and the M.S. and Ph.D. degrees from the University of California, Berkeley, in 1997 and 1999, respectively, all in electrical engineering.

He was with Hewlett-Packard Laboratories, Palo Alto, CA, from 1996 to 1998. After working as a Senior R&D Engineer in the Advanced Technology Group of Synopsys, Mountain View, CA, he is currently an Assistant Professor of Electrical Engineering at the University of Michigan, Ann Arbor. He has published numerous papers in his field of research, which includes the modeling, characterization, and analysis of on-chip interconnect, low-power circuit design techniques, and variability-aware circuit approaches.

Dr. Sylvester received an NSF CAREER award, the 2000 Beatrice Winner-Award at ISSCC, two outstanding research presentation awards from the Semiconductor Research Corporation, and a best student paper award at the 1997 International Semiconductor Device Research Symposium. He is also the recipient of the 2003 Ruth and Joel Spira Outstanding Teaching Award in the University of Michigan College of Engineering. His dissertation research was recognized with the 2000 David J. Sakrison Memorial Prize as the most outstanding research in the Electrical Engineering and Computer Science Department of the University of California, Berkeley. He is on the technical program committee of several design automation and circuit design conferences and was the general chair for the 2003 ACM/IEEE System-Level Interconnect Prediction (SLIP) Workshop. In addition, he is part of the International Technology Roadmap for Semiconductors (ITRS) U.S. Design Technology Working Group and made significant modeling contributions to the Design and System Drivers chapters of the 2001 ITRS. He is a Member of the Association for Computing Machinery, American Society of Engineering Education, and Eta Kappa Nu.