Analytical Yield Prediction Considering Leakage/Performance Correlation

Rajeev R. Rao, Anirudh Devgan, *Senior Member, IEEE,* David Blaauw, *Member, IEEE,* and Dennis Sylvester, *Senior Member, IEEE*

***Abstract*—In addition to traditional constraints on frequency, leakage current has emerged as a stringent constraint in modern processor designs. Since leakage current exhibits a strong inverse correlation with circuit delay, effective parametric yield prediction must consider the dependence of leakage current on frequency. In this paper, a new chip-level statistical method to estimate the total leakage current in the presence of within-die and die-to-die variability is presented. A closed-form equation for total chip leakage that models the dependence of the leakage current distribution on different process parameters is developed. The proposed analytical expression is obtained directly from pertinent design information and includes both subthreshold and gate leakage currents. Using this model, an integrated approach to accurately estimate the yield loss when both frequency and power limits are imposed on a design is then presented. The proposed method demonstrates the importance of considering both these limiting factors while calculating the yield of a lot.**

***Index Terms*—Estimation, leakage current, parametric yield, simulation.**

I. Introduction

CONTINUED scaling of device dimensions combined with shrinking threshold voltages has enabled designers to produce integrated circuits (ICs) that contain hundreds of millions of devices. However, this has also resulted in an exponential rise in IC power dissipation. This increase is primarily due to leakage, which is emerging as a significant portion of the total power consumption. It is estimated that the subthreshold leakage power will account for over 50% of the total power for portable applications developed for the 65-nm technology node [1]. In future technologies, aggressive scaling of the oxide thickness will lead to significant gate oxide tunneling current further aggravating the leakage problem. It is estimated that across successive technology generations, subthreshold leakage increases by about 3–5× [2], while gate leakage can increase by as much as 30× [3].

At the same time, the increased presence of parameter variability in modern designs has accentuated the need to consider the impact of statistical leakage current variations during the design process. For ±10% variation in the effective channel length of a transistor, there can be up to a 3× difference in the amount of subthreshold leakage current [4]. Gate leakage current exhibits great sensitivity to process variations showing a 6× difference in current for a ±4% variation in oxide thickness in the 100-nm Berkeley Predictive Technology Model (BPTM) [5]. Hence, considerable variability in chip-level leakage current can be expected and measured variations as high as 20× have been reported in literature [6].

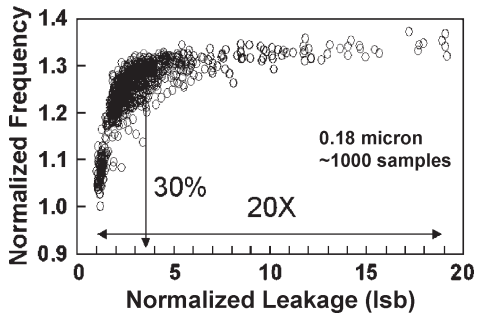


Fig. 1. Leakage and frequency variations (source: Intel).

In current designs, the parametric yield of a lot is typically calculated by characterizing the chips according to their operating frequency. Parametric yield refers to the ratio of chips that meet performance requirements to the total number of functional chips. The subset of dies that do not meet the required performance constraint are rejected, making this aspect of the design process very important from a commercial point of view. However, it has been observed [6] that among the "good" chips that meet the performance constraint, a substantial number of chips dissipate very large amounts of leakage power and, thus, are unsuitable for commercial usage. This incongruity is due to the inverse correlation between circuit delay and leakage current. Although the delay is reduced for devices with channel lengths smaller than the nominal value, it has the negative effect of vastly increasing the leakage current, resulting in higher leakage dissipation for chips with high operating frequencies.

The inverse correlation is illustrated in Fig. 1, which shows the distribution of chip performance and leakage based on silicon measurements over a large number of samples of a high-end processor design [6]. As can be seen, both the mean frequency and the variance of the leakage distribution increase significantly for chips with higher frequencies. This trend is particularly troubling, since it substantially reduces the yield of designs that are both performance and leakage constrained. Hence, there is a need for accurate leakage yield prediction methods that model this dependency.

Manuscript received September 24, 2004; revised March 2, 2005. This work was supported in part by the National Science Foundation, the Semiconductor Research Corporation (SRC), the Gigascale Silicon Research Center (GSRC)/Defense Advanced Research Projects Agency, IBM, and Intel. This paper was recommended by Associate Editor F. N. Najm.

R. R. Rao, D. Blaauw, and D. Sylvester are with the University of Michigan, Ann Arbor, MI 48109 USA (e-mail: rrrao@eecs.umich.edu; blaauw@eecs.umich.edu; dennis@eecs.umich.edu).

A. Devgan is with Magma Design Automation, Austin, TX 78759 USA (e-mail: devgan@magma-da.com).

Digital Object Identifier 10.1109/TCAD.2005.858351

Several statistical methods have been suggested to estimate the full-chip leakage current. In [7], the authors consider within-die threshold voltage variability to estimate the full-chip subthreshold leakage current. A compact current model is used in [8] to estimate the total leakage current. Rao *et al.* [9] present analytical equations to model subthreshold leakage as a function of the channel length of the transistor. A moment-based approximation approach is used to estimate the mean and variance of leakage current in [4] and [10]. However, none of these methods provide closed-form mathematical equations to express the chip leakage. Furthermore, they do not consider the dependence of leakage on frequency. More recently, a number of mathematical models have been proposed to determine the parametric yield of a lot. Najm and Menezes [11] use principal component analysis in order to estimate the timing yield. The circuit resizing algorithm proposed in [12] ensures the delay optimality of the circuit while achieving a desired yield figure. The probabilistic framework presented in [13] estimates the full-chip subthreshold leakage power distribution as well as the leakage constrained yield under the impact of process variations. In [14], the authors present a survey on the impact of technology scaling and process variations on the efficacy of various leakage reduction schemes.

In this paper, we develop a complete stochastic model for leakage current that includes the effects from multiple sources of variability and captures the dependence of the leakage current distribution on operating frequency. We consider the contribution from both interdie and intradie process variations and model total leakage as consisting of both subthreshold and gate tunneling leakage currents. We derive a closed-form expression for the total leakage as a function of all relevant process parameters. Next, we consider the impact of various process parameters on the chip frequency (performance). In our analysis, we assume that frequency is only affected by global variations in effective channel length and present recent industrial data to support this claim. We also present an analytical equation to quantify the yield loss when a power limit is imposed. This method precludes the need to use time-intensive circuit simulations to characterize the leakage current of a chip and enables the designer to budget for yield loss before the chip is sent to production. The proposed analytical expression is then compared with Monte Carlo simulation using SPICE measurements on a large circuit block to demonstrate its accuracy. Finally, we construct yield curves to accurately estimate the number of chips that satisfy both power and frequency constraints.

The remainder of this paper is organized as follows. In Section II, we present the model for full-chip total leakage. The models for subthreshold and gate leakages are presented separately. In Section III, we derive analytical equations to describe the yield prediction of a lot based on our leakage model. In Section IV, we present results, and in Section V, we conclude the paper.

II. Full-Chip Leakage Model

In this section, we present an analytical model to determine total leakage current expended by a chip. We model the leakage current as a function of different process parameters. First, we note that the total leakage is a sum of the subthreshold and gate leakages, written as

$$I_{\rm tot} = I_{\rm sub} + I_{\rm gate}. \tag{1}$$

Recently, it has been noted that other types of leakage current, such as band-to-band tunneling (BTBT), may become prominent in future process technologies [8]. Although we do not model other types of leakage in this paper, our analysis can be easily extended to include these additional leakage components.

In the subsequent sections, we model each type of leakage separately. We express both types of leakage current as a product of the nominal value and a multiplicative function that represents the deviation from the nominal due to process variability as

$$I_{\rm leakage} = I_{\rm nominal} \cdot f(\Delta P) \tag{2}$$

where P is the process parameter that affects the leakage current $I_{\rm leakage}$. In general, f is a nonlinear function. Since estimation methods based directly on Berkeley SPICE simulator (BSIM) models [15] are often overly complex, we use carefully chosen empirical equations in our analysis to provide both efficiency and accuracy.

We further decompose parameter P into two components as

$$\Delta P = \Delta P_{\rm global} + \Delta P_{\rm local} \tag{3}$$

where $\Delta P_{\rm global}$ models the global (die-to-die or interdie) process variations and $\Delta P_{\rm local}$ represents the local (within-die or intradie) process variations. For the sake of simplicity, we have assumed in our analysis that different types of variations at different levels (wafer, lot, etc.) in the chip manufacturing process are all contained in the die-to-die variation component. In a typical manufacturing process, both $\Delta P_{\rm global}$ and $\Delta P_{\rm local}$ are generally modeled as independent normal random variables (RVs) making ΔP also a normal RV. Since we are only dealing with the deviation from the nominal value, ΔP is a zero mean variable. If P is the effective channel length, then $\Delta P_{\rm local}$ is the term for the so-called across-chip linewidth variations (ACLVs). For simplicity of notation, we let $\Delta P_{\rm global} = P_g$ and $P_{\rm local} = P_l$ for all the parameters.

A. Subthreshold Leakage

Subthreshold leakage current $(I_{\rm sub})$ refers to the source-to-drain current when the transistor has been turned OFF. As is well known, $I_{\rm sub}$ has an exponential relationship with the threshold voltage $V_{\rm th}$ of the device, shown as

$$I_{\rm sub} = I_{\rm nominal} e^{[f(\Delta V_{\rm th})]}. \tag{4}$$

For the 0.13-μm technology node, even small variations in $V_{\rm th}$ can, therefore, result in leakage numbers that differ by a factor of $5-10\times$ from the nominal value.

Threshold voltage is a technology-dependent variable that must be expressed as a function of a number of parameters. The standard BSIM4 description [15] models several device

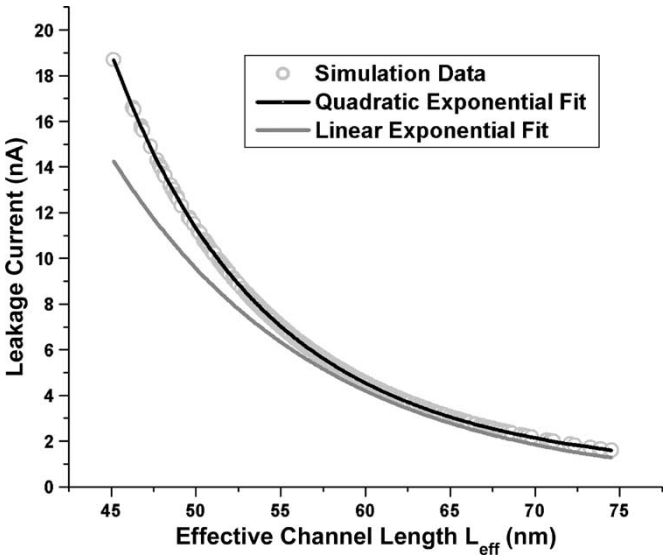


Fig. 2. Comparison of quadratic and linear exponential fit of leakage with effective channel length. Nominal $L_{\rm eff} = 60$ nm.

characteristics [short channel effect, drain-induced barrier lowering (DIBL), narrow-width effect, etc.] that influence $V_{\rm th}$ and expresses it as a function of several process parameters including effective channel length $(L_{\rm eff})$, doping concentration $(N_{\rm sub})$, and oxide thickness $(T_{\rm ox})$. Among these parameters, the variation in $L_{\rm eff}$ has the greatest impact as noted in [9]. A second-order but still significant portion of the variation in $V_{\rm th}$ occurs due to fluctuations in doping concentration that result in different values of the flat-band voltage $V_{\rm fb}$ for different transistors on the chip [4]. Finally, oxide thickness is a fairly well-controlled process parameter, and since the sensitivity of subthreshold leakage to oxide thickness is very small [4], [10], we do not include $T_{\rm ox}$ in the list of parameters that affect $I_{\rm sub}$. Hence, we empirically model the variation in $V_{\rm th}$ as an algebraic sum of two terms, expressed as

$$f(\Delta V_{\rm th}) = f[\Delta L_{\rm eff}] + f\left[\Delta(V_{\rm th,Nsub})\right] \tag{5}$$

where

$f(\Delta L_{\rm eff})$ variation in $V_{\rm th}$ due to variations in effective channel length of device;

$f(\Delta V_{\rm th,Nsub})$ variation in $V_{\rm th}$ due to doping concentration variations.

In our approach, we model $\Delta L_{\rm eff}$ and $\Delta V_{\rm th,Nsub}$ as independent normal RVs. Although there is a minor dependency between these two variables, we found the amount of error introduced by this independence assumption negligible.

Previously, leakage had been modeled as a single exponential function of the effective channel length [7], but a polynomial exponential model is much more accurate in capturing the dependency of leakage on effective channel length [9]. In Fig. 2, we plot both the quadratic and linear exponential fit for leakage current with respect to the effective channel length. In this plot, for the sake of simplicity, we set the variation in $V_{\rm th}$ due to doping to be equal to zero. We can clearly infer from this plot that since the linear exponential model underestimates the leakage current (especially at smaller channel lengths), it proves to be insufficient to model leakage current. Hence, we use a quadratic exponential model to express the variation in subthreshold leakage with respect to $L_{\rm eff}$. On the other hand, for $f(\Delta V_{\rm th,Nsub})$, we determined from circuit simulations that a linear model is sufficient. For simplicity of notation, let $\Delta L_{\rm eff} = L$ and $\Delta V_{\rm th,Nsub} = V$. Using this, we can rewrite (4) as

$$\begin{aligned} f(\Delta L_{\rm eff}) &= \frac{-(L + c_2 L^2)}{c_1} \\ f(\Delta V_{\rm th,Nsub}) &= -\left(\frac{c_3}{c_1}\right) V \\ I_{\rm sub} &= I_{\rm sub,nom} e^{-\left[\frac{L + c_2 L^2 + c_3 V}{c_1}\right]}. \end{aligned} \tag{6}$$

Here, c_1, c_2, and c_3 are fitting parameters and $I_{\rm sub,nom}$ is the subthreshold leakage of the device in the absence of any variability. The negative sign in the exponent is indicative of the fact that transistors with shorter channel lengths and lower threshold voltages produce higher leakage current. The empirical constants c_1, c_2, and c_3 can be determined by sweeping the effective channel length value over the required range and fitting the resultant subthreshold leakage current according to (6).

Using (3), we decompose L and V into local (L_l, V_l) and global (L_g, V_g) components, and we write the $I_{\rm sub}$ equation as

$$\begin{aligned} L &= L_g + L_l \\ V &= V_g + V_l \\ I_{\rm sub} &= I_{\rm sub,nom} e^{-\left[\frac{L_g + c_2 L_g^2 + c_3 V_g}{c_1}\right]} e^{-\left[\frac{L_l + \lambda_2 L_l^2 + \lambda_3 V_l}{\lambda_1}\right]} \end{aligned} \tag{7}$$

where $I_{\rm sub}$ is the subthreshold leakage of a single device with unit width. The mapping from c_i to λ_i $(i = 1, 2, 3)$ is obtained by writing $\lambda_1 : \lambda_2 : \lambda 3 = c_1 : c_2 : c_3$, where the constant of proportionality $\psi = \lambda_i / c_i$ is determined using the equation $\psi = 1/(1 + 2c_2 L_g)$. By definition L_l, which is the within-in die channel length variation, is a zero-mean RV. The within-die variation in a process parameter is typically composed of a systematic (correlated, layout dependent) component and a random (independent) component [16]. Rao *et al.* [9] show that for designs consisting of at least 250 individual clusters and a die area of 2.5 mm^2, the effect of spatial correlation on the total chip leakage current can be ignored. We use this result in our subsequent analyses and assume that the within-die variation in a process parameter is entirely due to the random (independent) component.

To calculate the total subthreshold leakage for a chip, we need to add the leakages device by device, considering that each device has unique RVs L_l and V_l, while sharing the same RVs L_g and V_g with all other devices. From (7), $I_{\rm sub} = g(L_l)$, where the subthreshold leakage distribution of a single transistor expressed as a function of L_l can be considered as an RV. The total subthreshold leakage is then a sum of all these individual RVs. If the number of RVs is large enough, the variance of their sum approaches zero. Consequently, we use the Central Limit theorem [17] to approximate the distribution of this sum by a single deterministic number. (We point out that the Central Limit theorem is applicable when the number

of individual RVs is 30 or more.) Further, if the number of RVs is large enough, the single number will be the mean of the distribution of this sum. Since modern complementary metal oxide semiconductor (CMOS) designs contain millions of devices that are distributed over a relatively large area on the chip, we use the independence assumption and substitute the sum of leakages over all devices with the mean value of $I_{\rm sub}$ over the complete range of L_l. This mean value is a simple scaling factor [7] that describes the relation between $I_{\rm sub}$ and L_l. Local variations are often spatially correlated, which means that devices that are positioned close together have positive correlation. However, as long as there are sufficient independent regions (clusters) on a die (as is typically the case), the Central Limit theorem can be applied. We use a similar method to calculate the scaling factor for the local variability of each process parameter.

To calculate the scaling factor, we need to find an exact expression for the expected value (mean) of $I_{\rm sub}$. Since $I_{\rm sub}$ is a function of two independent RVs, (V_l, L_l), we write a double integral to calculate the mean as

$$E[I_{\rm sub}] = \int_{-\infty}^{\infty}\left(\left[\int_{-\infty}^{\infty} g(L_l)\,{\rm pdf}(L_l)\,dL_l\right] g(V_l)\,{\rm pdf}(V_l)\,dV_l\right)$$
$$g(L_l) = I_{\rm sub,nom} e^{-\left[\frac{L_g + c_2 L_g^2}{c_1}\right]} e^{-\left[\frac{L_l + \lambda_2 L_l^2}{\lambda_1}\right]}$$
$$g(V_l) = e^{-\left[\frac{c_3 V_g^2}{c_1}\right]} e^{-\left[\frac{\lambda_3 V_l^2}{\lambda_1}\right]}. \tag{8}$$

In this equation, the terms containing L_g and V_g are constant for a given chip. The above integrals can be solved in closed form using the expressions given in (22) and (24) of the Appendix. We obtain $I_{\rm sub}$ as

$$I_{\rm sub} \approx E[I_{\rm sub}] = S_L S_V I_{L_g,V_g}$$
$$S_L = \left[\frac{1}{\left(\sqrt{1+\frac{2\lambda_2}{\lambda_1}\sigma_{L_l}^2}\right)}\right] e^{\left[\frac{\sigma_{L_l}^2}{\left(2\lambda_1^2 + 4\sigma_{L_l}^2 \lambda_1 \lambda_2\right)}\right]}$$
$$S_V = e^{\left(\frac{\lambda_3^2 \sigma_{V_l}^2}{2\lambda_1^2}\right)}$$
$$I_{L_g,V_g} = I_{\rm sub,nom} e^{-\left[\frac{L_g + c_2 L_g^2}{c_1}\right]} e^{-\left[\frac{c_3 V_g}{c_1}\right]} \tag{9}$$

where S_L and S_V are the scale factors introduced due to local variability in L and V I_{L_g,V_g} corresponds to the subthreshold leakage as a function of global variations.

Equation (9) provides the average value of subthreshold leakage for a unit width device. To compute the total chip subthreshold leakage, we need to perform a weighted sum of the leakages of all devices by considering the device widths to be the weights. For complex gates (transistor stacks, registers), a scale factor (q) model [7] is used to predict the effect of the total device width as

$$I_{c,\rm sub} = S_L S_V \left[\sum_d \left(\frac{WS_d}{q_s}\right) I_{L_g,V_g}\right]. \tag{10}$$

Here, the term $\sum_d (WS_d/q_s) I_{L_g,V_g}$ represents the chip-level subthreshold leakage as a function of the global process parameters (L_g, V_g). The WS_d term refers to the widths of the devices that contribute to the subthreshold leakage and the q_s term refers to the scaling factor for subthreshold leakage. We see from this equation that $I_{c,\rm sub}$ is a random variable rather than a single deterministic value.

B. Gate Leakage

When the oxide thickness of a device is reduced, there is an increase in the amount of carriers that can tunnel through the gate oxide. This phenomenon leads to the presence of gate leakage current $(I_{\rm gate})$ between the gate and substrate as well as the gate and channel. $I_{\rm gate}$ is linearly dependent on the area of the device and has an exponential relationship with the oxide thickness $(T_{\rm ox})$. Since the variation in $T_{\rm ox}$ has, by far, the greatest impact on gate leakage, we model $I_{\rm gate}$ as

$$I_{\rm gate} = I_{\rm gate,nom} e^{[f(\Delta T_{\rm ox})]}. \tag{11}$$

From circuit simulation, we found that it is sufficient to express $f(\Delta T_{\rm ox})$ as a simple linear function. A suitable value for a single parameter β_1 efficiently captures the highly exponential relationship. Let $\Delta T_{\rm ox} = T$. Using (3), we decompose T into global (T_g) and local (T_l) components as

$$f(\Delta T_{\rm ox}) = -\left(\frac{T}{\beta_1}\right)$$
$$T = T_g + T_l$$
$$I_{\rm gate} = I_{\rm gate,nom} e^{-\left(\frac{T_g}{\beta_1}\right)} e^{-\left(\frac{T_l}{\beta_1}\right)}. \tag{12}$$

$I_{\rm gate}$ is the gate leakage current of a single device with unit width. The empirical constant β_1 can be determined by sweeping the oxide thickness value over the required range and fitting the resultant gate leakage current according to (12). $I_{\rm gate,nom}$ is the nominal gate leakage and both T_g and T_l are zero-mean RVs. We see that the relationship between $I_{\rm gate}$ and T_l is similar to the single exponential relationship between $I_{\rm sub}$ and V_l. Similar to S_V, we compute the scale factor S_T due to T_l as

$$I_{\rm gate} \approx E[I_{\rm gate}] = S_T I_{T_g}$$
$$S_T = e^{\left(\frac{\sigma_{T_l}^2}{2\beta_1^2}\right)}$$
$$I_{T_g} = I_{\rm gate,nom} e^{-\left(\frac{T_g}{\beta_1}\right)}. \tag{13}$$

Based on the widths of the devices, the chip-level gate leakage can be calculated in a similar manner as the subthreshold leakage as

$$I_{c,\rm gate} = S_T \left[\sum_d \left(\frac{WG_d}{q_g}\right) I_{T_g}\right] \tag{14}$$

where WG_d term refers to the widths of the devices that contribute to the gate leakage and the q_g term refers to the scaling factor for gate leakage.

C. Total Leakage

The total leakage is the sum of the subthreshold and gate leakage currents of all the devices. In (10) and (14), we note that in I_{L_g,V_g} and I_{T_g}, the global variability components are shared by all the devices on the chip. Hence, we can write the equation for total chip leakage as

$$I_{c,\text{tot}} = S_L S_V \left[\sum_d \left(\frac{WS_d}{q_s}\right) I_{L_g,V_g}\right] + S_T \left[\sum_d \left(\frac{WG_d}{q_g}\right) I_{T_g}\right]. \quad (15)$$

This equation can be used to calculate the total leakage for different types of devices such as n-channel metal oxide semiconductor (NMOS)/p-channel metal oxide semiconductor (PMOS) and low/high-V_{th} transistors. The differences will be in the fitting parameters and the scale factors q_s and q_g. The sum total over all devices gives the total leakage of the chip.

III. Yield Analysis

Traditional parametric yield analysis of high-performance ICs is done using the frequency (or speed) binning method [18]. For a given lot, each chip is characterized according to its operating frequency and figuratively placed in a particular bin according to this value. A frequency limit is specified and chips that operate at frequencies below this limit are discarded. As illustrated in Fig. 1, due to the inverse correlation between leakage and circuit delay, chips in the "fast" corner produce vast amounts of leakage current compared to the other chips. In current technologies, this is a major concern, since a significant number of these chips leak more than the acceptable value and must be discarded [19]. Thus, parametric yield loss is exacerbated as a result of dies now being lost at both the low- and high-speed bins, further narrowing the acceptable process window. The binning process has a significant impact on product profitability.

In this section, we describe a method to calculate the yield of a lot when both frequency and power limits are imposed. We first show that chip frequency is most strongly influenced by global gate length variability and, hence, as in standard industry practice, each frequency bin corresponds to a specific value of L_g. We then compute the yield due to the imposed leakage limit on a bin-by-bin basis.

A. Frequency Dependence on Process Parameters

In principle, chip frequency depends on a number of process parameters. From the model presented in this paper, we see that the chip frequency is affected by six main parameters—three global and three local variability components of effective channel length (L), doping concentration (V), and oxide thickness (T). We first demonstrate from SPICE simulations that circuit delay is primarily impacted by gate length variations. For this purpose, we simulated a 17-stage ring oscillator under different process conditions using the 100-nm BPTM process technology as shown in Fig. 3. In this plot, we sweep the global variability components L_g, V_g, and T_g on the x axis and calculate the delay of the ring oscillator for a range of parameter values. From this plot, we see that variations in L_g significantly influence (by approximately $\pm 15\%$) the delay of the ring oscillator. Variations in V_g and T_g have little or no impact on the delay and can, thus, be ignored. By a similar argument, it follows that variations in the local variability components V_l and T_l also have negligible impact on circuit delay and can be neglected during chip performance analysis.

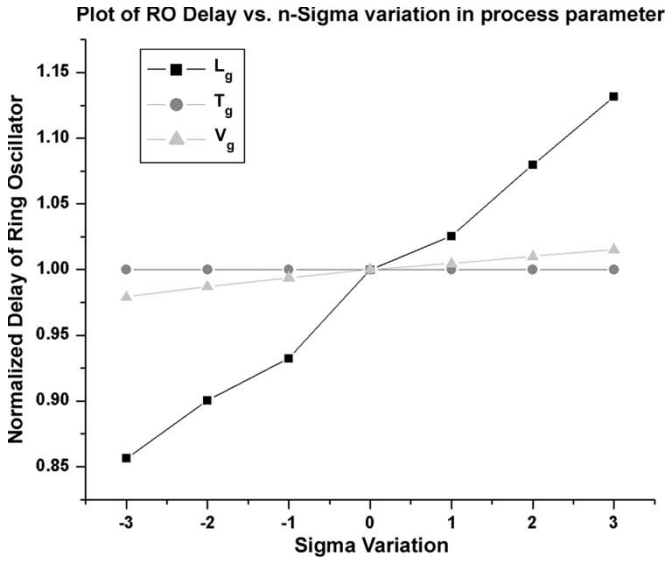


Fig. 3. Comparison of relative contribution of parameter variations on ring oscillator delay.

From this analysis, we have determined that only the global (L_g) and local (L_l) variability components of effective channel length affect the chip frequency value. It has been pointed out previously in literature [20] that within-die variations primarily impact the mean of the maximum frequency (FMAX) while die-to-die variations affect the variance of the distribution in maximum frequency. Thus, both $L_g$and L_l determine the value of chip frequency. In our subsequent analysis, we do not model the effect of L_l on chip performance and instead assume that frequency is only dependent on the value of L_g. While this is a simplifying approximation, we point to recent experimental data from industry showing that the impact of within-die channel length variability (L_l) on chip performance is minimal [23], [24].

In order to support this approximation, we first observe that the number of critical paths in a design plays a critical role in measuring the impact of variability on the maximum frequency of a chip. In general, for a given circuit path, the effect of local variability tends to average out over the various devices in the path and, thus, the impact of local variations on circuit delay is insignificant compared to global variations [21]. Therefore, if the circuit contains a single critical path, the effect of local variability is minimal. As the number of critical paths increases, the impact of local variability increases dramatically, because, due to their spatial location and composition, certain circuit paths can become slower or faster, thereby affecting the entire

chip frequency distribution. Orshansky *et al.* [22] highlight this important fact and present an analytical model where the mean chip frequency can be shifted by as much as 2σ when there are 100 critical paths. When the number of critical paths increases, a large number of circuit path delays are close to the chip maximum frequency and form a so-called timing wall [22]. A larger number of paths close to the timing wall is indicative of a circuit that has been optimized through efficient pipelining and logic balancing.

However, it has been recently pointed out [23] that even for a well-debugged product in which 50% of all critical paths are within 2.5% of the worst path delay, the impact of within-die variation on chip maximum frequency is very small. Samaan [23] uses the analytical model from [24] on industrial circuits to show that for within-die variation with $3\sigma = \pm 10\%$, the FMAX value is no worse than 5% of its original value. Even if the 3σ value of within-die variation and the number of critical paths increases drastically, the reduction in FMAX value is at most 9%. The author asserts that this result on critical paths, coupled with the fact that within-die variation will be kept under relative control (according to the ITRS [25]), indicates that within-die variation will not affect the chip performance by more than a few-to-several percent for at least the next four technology generations. The author concludes that random within-die variation will have a small and manageable impact on chip frequency until the end of this decade.

Hence, although local variations affect the circuit delay, we choose to ignore their effect in our yield computation. The results from the subsequent analysis can be viewed as normalizing the global (die-to-die) variability to the degraded mean chip delay that results from the within-die variations. We assume that there is a one-to-one correspondence between chip frequency (performance) and the global effective channel length variability value (L_g). This is consistent with current industry practices, where a direct correspondence is often assumed between frequency bins and specific gate length values. The impact of local variability on-chip frequency will be considered in future extensions of our work.

B. Yield Estimate Computation

We now discuss the method to compute the expected yield for a particular frequency bin based on an imposed leakage limit. For a particular bin, the value of L_g is available, and using the expressions for I_{L_g,V_g} (9) and I_{T_g} (13), we rewrite the equation for total chip leakage (15) as

$$\begin{aligned} I_{\text{tot}} &= A_s e^{\left(\frac{V_g}{k_v}\right)} + A_g e^{\left(\frac{T_g}{k_t}\right)} \\ A_s &= \left(\sum_d \left(\frac{WS_d}{q_s}\right)\right) S_L S_V I_{\text{sub,nom}} e^{-\left[\frac{\left(L_g + c_2 L_g^2\right)}{c_1}\right]} \\ A_g &= \left(\sum_d \left(\frac{WG_d}{q_g}\right)\right) S_T I_{\text{gate,nom}} \\ k_v &= -\left(\frac{c_1}{c_3}\right) \\ k_t &= -\beta_1. \end{aligned} \tag{16}$$

Here, we simplified the notation for the fitting parameters and expressed this equation in terms of the new constants k_v and k_t. The values for k_v and k_t are generally expressed in terms of σ_{V_g} and σ_{T_g}. A_s represents the total chip subthreshold leakage at a particular value of L_g and includes the scale factors due to the local variability. Similarly, A_g represents the total chip gate leakage at a given value of L_g. However, since $I_{c,\text{gate}}$ is independent of L_g, A_g is not influenced by changes in the value of L_g. In a plot of total leakage versus L_g, we first compute A_s and A_g at particular values of L_g and then calculate the distribution of I_{tot} at each of these points.

For every device type, I_{tot} is the sum of two lognormal variables, one of which represents the subthreshold leakage current and the other represents the gate leakage current. For a particular device, by our formulation, there is no parameter that affects both of these terms simultaneously. Therefore, we can consider these terms as independent RVs. For a given circuit design, the total leakage will then be the sum of a small set of lognormals with each device type contributing exactly two lognormals to the total leakage set. We use Wilkinson's method [26] to model this sum of lognormals as another lognormal RV. Using the independence condition, we set the sums of the means and variances to be equal to the mean and variance of the new lognormal. From (22) in the Appendix, we get (17) as specified here. While (17) has been developed contingent on the independence of X_1 (subthreshold) and X_2 (gate), we observe that the dependency condition can be easily incorporated into our analysis by including extra terms in (17) to account for the correlation between X_1 and X_2

$$\begin{aligned} I_{\text{tot}} &= X_1 + X_2 \\ X_1 &\sim \text{LN}\left(\log(A_s), \left(\frac{\sigma_{V_g}}{k_v}\right)^2\right) \\ X_2 &\sim \text{LN}\left(\log(A_g), \left(\frac{\sigma_{T_g}}{k_t}\right)^2\right) \\ \mu_{I_{\text{tot}}} &= \exp\left[\log(A_s) + \frac{1}{2}\left(\frac{\sigma_{V_g}}{k_v}\right)^2\right] \\ &\quad + \exp\left[\log(A_g) + \frac{1}{2}\left(\frac{\sigma_{T_g}}{k_t}\right)^2\right] \\ \sigma^2_{I_{\text{tot}}} &= \exp\left[2\log(A_s) + \left(\frac{\sigma_{V_g}}{k_v}\right)^2\right]\left[\exp\left(\frac{\sigma^2_{V_g}}{k_v^2}\right) - 1\right] \\ &\quad + \exp\left[2\log(A_g) + \left(\frac{\sigma_{T_g}}{k_t}\right)^2\right]\left[\exp\left(\frac{\sigma^2_{T_g}}{k_t^2}\right) - 1\right]. \end{aligned} \tag{17}$$

Equation (23) in the Appendix is then used to obtain the mean and variance $(\mu_{\text{N},I_{\text{tot}}}, \sigma^2_{\text{N},I_{\text{tot}}})$ of the normal RV corresponding to this lognormal. From these values, we can express the

TABLE I
VALUE OF $I_{\rm tot}$ FOR n-SIGMA POINT

n	$F_x(I_{tot})$	I_{tot}
0	0.500	$\exp(\mu_{\rm N,Itot})$
1	0.682	$\exp(\mu_{\rm N,Itot} + 0.473\sigma_{\rm N,Itot})$
2	0.954	$\exp(\mu_{\rm N,Itot} + 1.685\sigma_{\rm N,Itot})$
3	0.998	$\exp(\mu_{\rm N,Itot} + 2.878\sigma_{\rm N,Itot})$

TABLE II
COMPARISON OF SPICE SIMULATION AND ANALYTICAL DATA

Cases	Parameter sigma (σ) values			Mean Leakage (μA)	
	(L_g, L_l)	(V_g, V_l)	(T_g, T_l)	Exp	Ana
No variation	(0, 0)	(0, 0)	(0, 0)	14.97	15.22
Only die-to-die	(-1, 0)	(-1, 0)	(-1, 0)	20.82	21.32
Both variations	(-1, ±3)	(-1, ±3)	(-1, ±3)	24.01	24.95

probability density function (pdf) of the total leakage using the standard expression for the pdf of a lognormal RV as

$$\mu_{{\rm N},I_{\rm tot}} = \frac{1}{2}\log\left[\frac{\mu_{I_{\rm tot}}^4}{\left(\mu_{I_{\rm tot}}^2 + \sigma_{I_{\rm tot}}^2\right)}\right]$$

$$\sigma_{{\rm N},I_{\rm tot}}^2 = \log\left[1 + \left(\frac{\sigma_{I_{\rm tot}}^2}{\mu_{I_{\rm tot}}^2}\right)\right]$$

$$\begin{aligned}{\rm pdf}(I_{\rm tot}) = {} & \frac{1}{I_{\rm tot}\sqrt{2\pi\sigma_{{\rm N},I_{\rm tot}}^2}} \\ & \cdot \exp\left[-\left(\frac{\log(I_{\rm tot}) - \mu_{{\rm N},I_{\rm tot}}}{\sqrt{2}\sigma_{{\rm N},I_{\rm tot}}}\right)^2\right]. \quad (18)\end{aligned}$$

Finally, to obtain exact yield estimates, we require the quantile numbers for the lognormal distribution described by $I_{\rm tot}$ (i.e., the confidence points of $I_{\rm tot}$ that correspond to the specified leakage limit). Since the exponential function that relates $\mathrm{LN}(\mu_{I_{\rm tot}}, \sigma_{I_{\rm tot}}^2)$ with $\mathrm{N}(\mu_{{\rm N},I_{\rm tot}}, \sigma_{{\rm N},I_{\rm tot}}^2)$ is a monotonically increasing function, the quantiles of the normal RV are mapped directly to the quantiles of the lognormal RV [27]. Using this fact, we can write the expression for the cumulative distribution function (cdf) of a lognormal variable as

$$\begin{aligned}{\rm cdf}(I_{\rm tot}) &= F_x(I_{\rm tot}) \\ &= \frac{1}{2}\left[1 + {\rm erf}\left(\frac{\log(I_{\rm tot}) - \mu_{{\rm N},I_{\rm tot}}}{\sqrt{2}\sigma_{{\rm N},I_{\rm tot}}}\right)\right]. \quad (19)\end{aligned}$$

Here, erf() is the error function. By setting $F_x(\cdot)$ to a particular confidence point on the normal distribution, we can obtain the corresponding value on the lognormal distribution (see Table I). In Table I, the 0-sigma point corresponds to the median of the distribution.

Conversely, if we are given a limit for $I_{\rm tot}$, we can use (19) to compute $\mathrm{cdf}(I_{\rm tot})$ and determine the number of chips that meet the leakage limit in a particular performance bin. In a given frequency bin and for a given leakage limit, $[\mathrm{cdf}(I_{\rm tot}) \times 100]\%$ is the fraction of chips that meet both the speed and power criteria. Hence, by repeating this computation for each frequency bin that meets the frequency specification, the total percentage of chips that meet both the leakage and performance constraints can be found.

IV. RESULTS

In this section, we use our analytical method from the previous section to predict the yield of a lot. Our circuit of choice is a 64-bit adder written for the Alpha architecture. We assume that all dies in the lot consist of this circuit and a small ring oscillator circuit is used to characterize the frequency of the chip with the variation in L_g. We use the 100-nm ($L_{\rm eff} = 60$ nm) BPTM [5] for our SPICE Monte Carlo simulations. We also employ a gate leakage model based on the BSIM4 equations [15]. The variability numbers for $\Delta L_{\rm eff}$, $\Delta V_{\rm th,Nsub}$, and $\Delta T_{\rm ox}$ are based on estimates obtained from a typical industrial 90-nm process. The approximate values for the ratio $(3\sigma_{\rm tot}/\mu_{\rm tot})$ for L, V, and T were 15%, 10%, and 4%, respectively.

We first present a quantitative comparison between SPICE data and our analytical method. In Table II we consider three cases, namely: 1) no variability in any parameter; 2) only die-to-die variability; and 3) both within-die and die-to-die variability in all three parameters. The middle three columns correspond to the sigma variation values corresponding to each parameter. Thus, for the case when both types of variability are present, the global variability values for all three parameters are set to -1σ from the nominal while the local variability is set to be $\pm 3\sigma$ (this is the third row in Table II). We see from this table that for all the cases, the difference between the Monte Carlo simulation data and the analytical expressions is less than 5%. Furthermore, we note that the presence of local variability increases the amount of total chip leakage by about 15%.

Fig. 4 gives the scatter plot for 2000 samples of the total circuit leakage generated using SPICE. The y axis in the plot has been normalized to the sample mean of the leakage currents. We see that for the entire $\pm 3\sigma$ variation in L_g, there is a $14\times$ spread in the leakage. Additionally, for a given L_g, there is a wide "local" distribution in leakage. For instance, given $L_g = 0\sigma$, the normalized value of the total circuit leakage is between 0.5 and 1.7. In (16), we observe that even for small values of (V/k_v) and (T/k_t), the exponential terms increase rapidly and contribute a larger portion to the total leakage value. As a result, the distribution in V_g and T_g (for each value of L_g) produces a band-like curve for the scatter plot of total circuit leakage (instead of a single curve). This is significant, because for a given value of L_g (and, hence, a given operating frequency), a large portion of the chips may be about $3\times$ the nominal leakage value. A chip that operates at an acceptable frequency may still have to be discarded, because the variability in V_g and T_g pushes its leakage consumption over the tolerable limit. Thus, we see that the secondary variations V_g and T_g play a significant role in determining the yield of a lot.

In Fig. 5, we superimpose the analytically computed sigma contour lines on top of the same leakage scatter plot. For each value of L_g, we calculate $(\mu_{{\rm N},I_{\rm tot}}, \sigma_{{\rm N},I_{\rm tot}})$ and then use Table I to construct the contour lines. From the plot, we see that a fair number of samples lie "outside" the 1σ range. This is especially

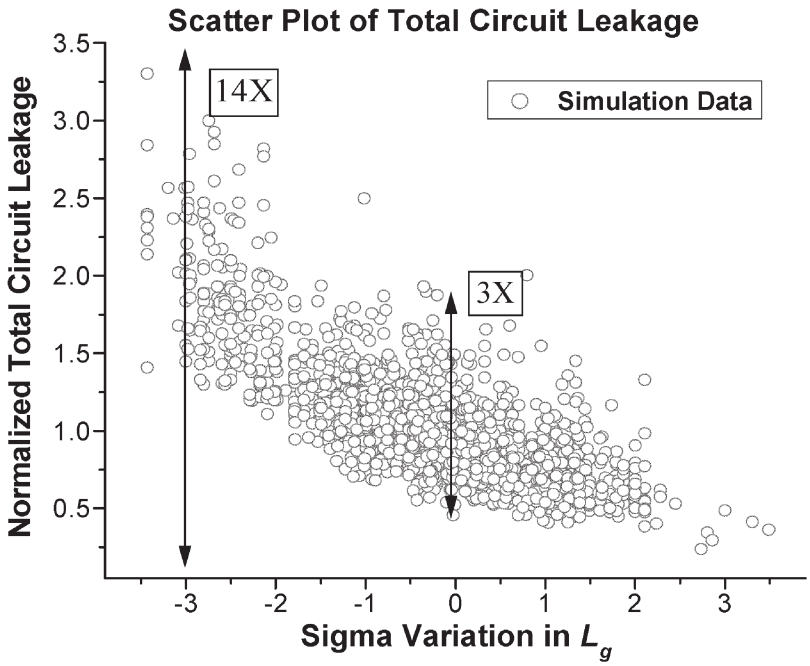


Fig. 4. Scatter plot showing distribution of total circuit leakage.

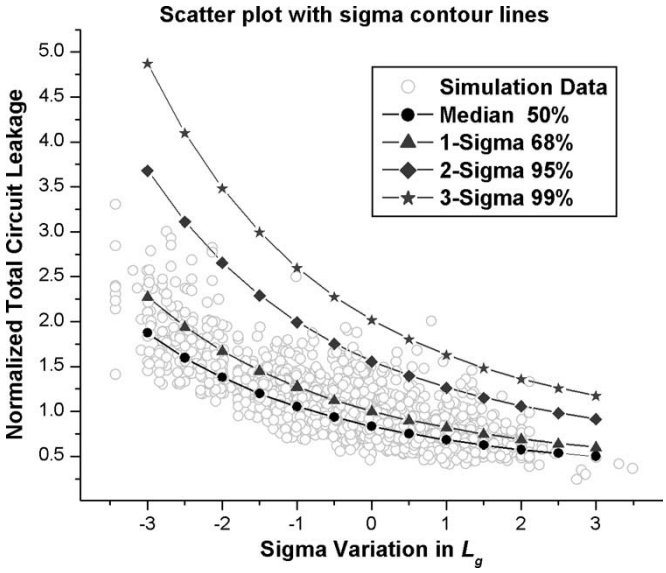


Fig. 5. Scatter plot of total circuit leakage with sigma contour lines added.

true for gate lengths close to the nominal. For shorter channel lengths, since the contour value is quite large, there are only a small number of chips outside this range. For larger channel lengths, since the absolute value of the leakage is quite small, there are practically no chips outside the 2σ range.

We now present an example calculation for the yield. For the lot presented here, we impose a frequency limit that corresponds to the frequency value of a chip at $L_g = +1\sigma$ and a normalized power limit $(P_{\lim})$ of 1.75 as indicated in Fig. 6. (Recall from Section III-A that based on our assumption, there exists a one-to-one correspondence between the chip frequency value and value of global channel length variability L_g.) The frequency bins are specified to be at the L_g n-sigma boundaries with $f_{(L_g=-3\sigma)} > f_{(L_g=-2\sigma)} > \cdots > f_{(L_g=+1\sigma)}$. All chips with frequencies in the range $[f_{(L_g=i\sigma)}, f_{(L_g=(i-1)\sigma)}]$ are assigned to the bin characterized by frequency $f_{(L_g=i\sigma)}$. This is a conservative estimate, since all the devices in a bin are assumed to be operating at the minimum possible frequency value for that bin.

First, we see that due to the performance (frequency) limit, all chips that operate at frequencies smaller than the $f_{(L_g=+1\sigma)}$ value are discarded. As we can see from the plot, although all of these chips meet the power criterion, they are discarded, because they are "too slow." Next, we proceed bin by bin and calculate the yield for each bin. To illustrate the yield computation, we present the numbers for the cases when the frequencies are specified at $L_g = [-3\sigma, -2\sigma, \ldots, +1\sigma]$. For each such L_g, we calculate the cdf values using (16)–(19). Table III summarizes these cdf numbers for $P_{\lim} = 1.75$. Fig. 7 presents the yield curves for seven different values of $P_{\lim}$. (In this figure, the line with $P_{\lim} = 1.75$ exactly corresponds to the values listed in Table III.) We note that this plot is essentially a discrete cdf curve describing the yield of a lot. The bin-by-bin analysis presented here mimics a similar operation done in the testing industry [18]. The use of binning is common in the chip industry and is dictated by commercial concerns, because chips with smaller values of L_g (and higher frequencies) are much more profitable than chips with higher values of L_g.

Traditional parametric yield analysis does not consider power as a criterion and, hence, overestimates the number of chips that are actually good/marketable. For instance, if $P_{\lim} = 1.75$, we see in Table III that for a frequency specified at $L_g = -2\sigma$, only 72.6% of the chips will meet the power criterion. Thus, even if the chip designer budgets for 1.75 times the nominal power, there is a loss of 27.4% of the chips operating in the fast corner. Furthermore, even for the nominal value of frequency at $L_g = 0\sigma$, about 2.5% of the chips are

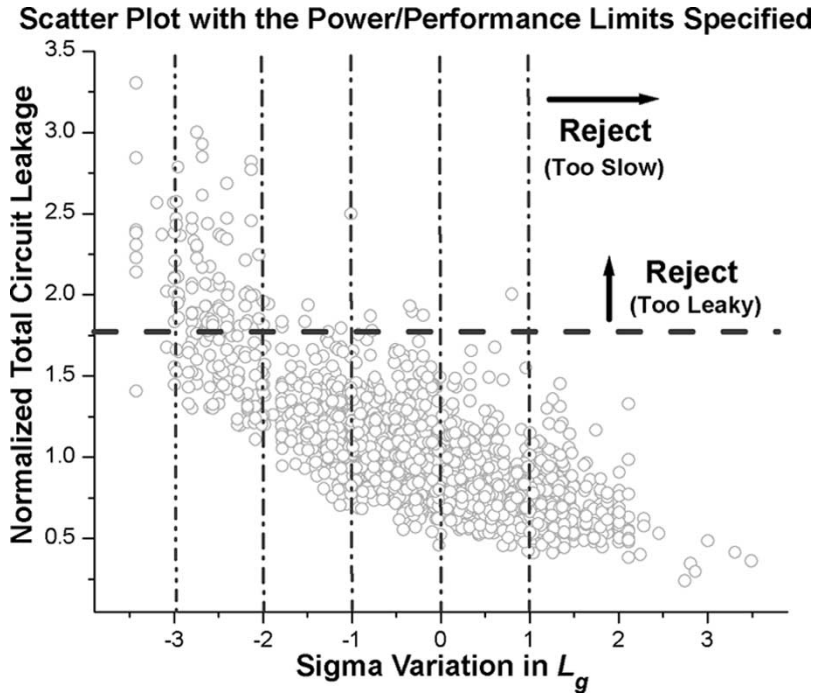


Fig. 6. Scatter plot with power and performance limits specified.

TABLE III
CDF$[(I_{\rm tot}) \times 100]\%$ NUMBERS FOR THREE DIFFERENT VALUES OF $P_{\rm lim}$ FOR OUR RANGE OF L_g VALUES

P_{lim}	L_g *n*-sigma				
	-3	-2	-1	0	1
1.75	43.6	72.6	90.5	97.5	99.4

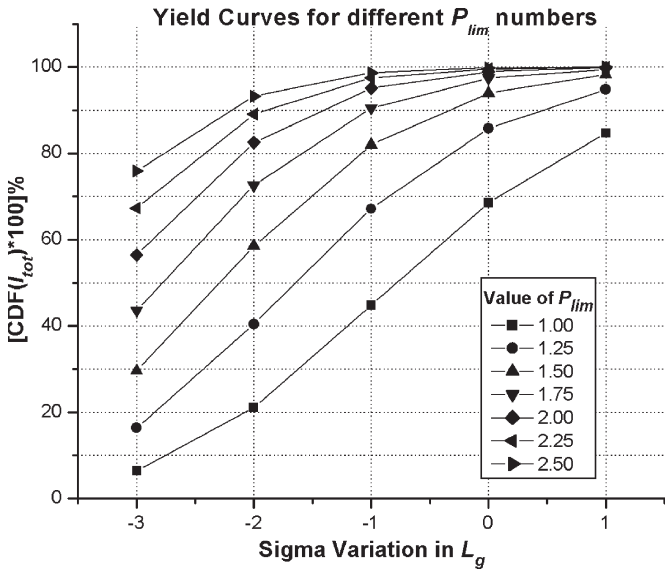


Fig. 7. Yield curves for different values of $P_{\rm lim}$.

lost, since they lie outside the power limit. While a typical frequency binning method would predict that 100% of the chips with frequency values at $L_g = -2\sigma$ are good, our method captures the fact that over 25% cannot be marketed. This is particularly important, since fast bin devices are considered to be highly profitable. Hence, there is a need to adopt an integrated approach that accounts for the compounded loss due to both limiting factors. We find that our approach always predicts a lower yield percentage compared to the method that assumes independence of the limiting factors of power and performance. By preserving the correlation between frequency and leakage, we are able to obtain more accurate estimates for the yield.

V. CONCLUSION

In this paper, we presented an analytical framework that provides a closed-form expression for the total chip leakage current as a function of relevant process parameters. We observed that the presence of local variability increases the total chip leakage by about 15%. Using our analytical expression, we estimated the yield of a lot when both power and performance constraints are imposed. We presented an example calculation for yield that shows the compounded loss that occurs due to chips that operate at low frequencies as well as chips that produce excessive amounts of leakage. Our method exemplifies the need to consider both limiters when calculating the yield of a lot.

APPENDIX

Given a normal (Gaussian) RV $X \sim \mathrm{N}(\mu_x, \sigma_x^2)$, the pdf of X is given by [17]

$$\mathrm{pdf}(x) = f_X(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\left(\frac{x-\mu_x}{\sqrt{2}\sigma_x}\right)^2\right]. \quad (20)$$

Given an RV X and a function $g(X)$, we form the new RV $Y = g(X)$ and express its mean and variance directly in terms

of the pdf of X as

$$E[g(X)] = \int_{-\infty}^{\infty} g(X)\,\text{pdf}(x)\,dx$$
$$E\left[\{g(X)\}^2\right] = \int_{-\infty}^{\infty} \{g(X)\}^2\,\text{pdf}(x)\,dx$$
$$\text{Var}[g(X)] = E\left[\{g(X)\}^2\right] - \{E[g(X)]\}^2. \quad (21)$$

The function $Y = g(X) = e^{-(X/a_1)}$ is a lognormal RV when X is a normal RV with $X \sim \text{N}(\mu_x, \sigma_x^2)$. The lognormal distribution is a continuous distribution where the natural logarithm of the variable has a normal distribution. (This is readily evident from the definition of $g(X)$, because $\ln(g(X)) = (-X/a_1)$ is obviously a normal RV.) The lognormal distribution is mainly used in reliability analysis to model failure times. It is also used to model diverse physical phenomena such as survival time of bacteria in disinfectants, the weight and blood pressure of humans, and the size of silver particles in a photographic emulsion [28].

Using the values for (μ_x, σ_x^2), we can express the mean and variance of Y in closed form. The integrals in (21) corresponding to this choice of $g(X)$ have been evaluated using the mathematical software package Maple [29] as

$$\mu_y = e^{\left[-\left(\frac{2\mu_x}{a_1}\right)+\left(\frac{\sigma_x^2}{2a_1^2}\right)\right]}$$
$$\sigma_y^2 = e^{\left[-\left(\frac{2\mu_x}{a_1}\right)+\left(\frac{\sigma_x^2}{a_1^2}\right)\right]}\left[e^{\left(\frac{\sigma_x^2}{a_1^2}\right)} - 1\right]. \quad (22)$$

Conversely, given the values for the mean and variance of the lognormal RV (μ_y, σ_y^2) as in (22), we can compute the mean and variance of the corresponding normal RV to obtain (μ_x, σ_x^2) (we have normalized Y by setting $a_1 = -1$) as

$$\mu_x = \frac{1}{2}\log\left[\frac{\mu_y^4}{\left(\mu_y^2 + \sigma_y^2\right)}\right]$$
$$\sigma_x^2 = \log\left[1 + \left(\frac{\sigma_y^2}{\mu_y^2}\right)\right]. \quad (23)$$

For the RV $Z = h(X) = e^{-[(X+a_2X^2)/a_1]}$, where X is a zero-mean normal RV, it is possible to obtain a closed-form expression for the mean and variance. The integrals in (21) corresponding to this value of $h(X)$ have again been evaluated, using the mathematical software package Maple, as

$$E[Z] = \mu_z = \left[\frac{1}{\left(\sqrt{1 + \frac{2a_2}{a_1}\sigma_x^2}\right)}\right] e^{\left[\frac{\sigma_x^2}{\left(2a_1^2 + 4\sigma_x^2 a_1 a_2\right)}\right]}$$
$$E[Z^2] = \left[\frac{1}{\left(\sqrt{1 + \frac{4a_2}{a_1}\sigma_x^2}\right)}\right] e^{\left[\frac{\sigma_x^2}{\left(\frac{a_1^2}{2} + 2\sigma_x^2 a_1 a_2\right)}\right]}$$
$$\sigma_z^2 = E[Z^2] - \mu_z^2. \quad (24)$$

ACKNOWLEDGMENT

The authors would like to thank Vivek De of Intel for providing them with Fig. 1. They also thank the anonymous reviewers for useful suggestions in improving this paper.

REFERENCES

[1] J. Kao, S. Narendra, and A. Chandrakasan, "Subthreshold leakage modeling and reduction techniques," in *Proc. ICCAD*, San Jose, CA, Nov. 2002, pp. 141–148.
[2] S. Borkar, "Design challenges of technology scaling," *IEEE Micro*, vol. 19, no. 4, pp. 23–29, Jul./Aug. 1999.
[3] K. Bernstein, C.-T. Chuang, R. Joshi, and R. Puri, "Design and CAD challenges in sub-90 nm CMOS technologies," in *Proc. ICCAD*, San Jose, CA, Nov. 2003, pp. 129–136.
[4] S. Mukhopadhyay and K. Roy, "Modeling and estimation of total leakage current in nano-scaled CMOS devices considering the effect of parameter variation," in *Proc. ISLPED*, Seoul, Korea, Aug. 2003, pp. 172–175.
[5] *Berkeley Predictive Technology Model (BPTM)*. [Online]. Available: http://www-device.eecs.berkeley.edu/~ptm/
[6] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," in *Proc. IEEE/ACM DAC*, Anaheim, CA, Jun. 2003, pp. 338–342.
[7] S. Narendra, V. De, S. Borkar, D. Antoniadis, and A. Chandrakasan, "Full-chip subthreshold leakage power prediction model for sub-0.1 μm CMOS," in *Proc. ISLPED*, Monterey, CA, Aug. 2002, pp. 19–23.
[8] S. Mukhopadhyay, A. Raychowdhury, and K. Roy, "Accurate estimate of total leakage current in scaled CMOS circuits based on compact current modeling," in *Proc. IEEE/ACM DAC*, Anaheim, CA, Jun. 2003, pp. 169–174.
[9] R. Rao, A. Srivastava, D. Blaauw, and D. Sylvester, "Statistical analysis of subthreshold leakage current for CMOS circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 2, pp. 131–139, Feb. 2004.
[10] A. Srivastava, R. Bai, D. Blaauw, and D. Sylvester, "Modeling and analysis of leakage power considering within-die process variations," in *Proc. ISLPED*, Monterey, CA, Aug. 2002, pp. 64–67.
[11] F. Najm and N. Menezes, "Statistical timing analysis based on a timing yield model," in *Proc. IEEE/ACM DAC*, San Diego, CA, Jun. 2004, pp. 460–465.
[12] S. Choi, B. Paul, and K. Roy, "Novel sizing algorithm for yield improvement under process variation in nanometer technology," in *Proc. IEEE/ACM DAC*, San Diego, CA, Jun. 2004, pp. 454–459.
[13] S. Zhang, V. Wason, and K. Banerjee, "A probabilistic framework to estimate full-chip subthreshold leakage power distribution considering within-die and die-to-die P–T–V variations," in *Proc. ISLPED*, Newport Beach, CA, Aug. 2004, pp. 156–161.
[14] Y.-F. Tsai, D. Duarte, N. Vijaykrishnan, and M. J. Irwin, "Impact of process scaling on the efficacy of leakage reduction schemes," in *Int. Conf. Integrated Circuit Design Technology (ICICDT)*, Austin, TX, May 2004, pp. 3–11.
[15] *BSIM4*. [Online]. Available: http://www-device.eecs.berkeley.edu/~bsim3/bsim4.html
[16] S. Nassif, "Modeling and forecasting of manufacturing variations," in *Int. Workshop Statistical Metrology*, Honolulu, HI, Jun. 2000, pp. 2–10.
[17] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1991.
[18] B. Cory, R. Kapur, and B. Underwood, "Speed binning with path delay test in 150-nm technology," *IEEE Des. Test Comput.*, vol. 20, no. 5, pp. 41–45, Oct. 2003.
[19] A. Keshavarzi, K. Roy, C. Hawkins, and V. De, "Multiple-parameter CMOS IC testing with increased sensitivity for I_{DDQ}," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 11, no. 5, pp. 863–870, Oct. 2003.
[20] K. Bowman, S. Duvall, and J. 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, no. 2, pp. 183–190, Feb. 2002.
[21] A. Agarwal, D. Blaauw, V. Zolotov, and S. Vrudhula, "Computation and refinement of statistical bounds on circuit delay," in *Proc. IEEE/ACM DAC*, Anaheim, CA, Jun. 2003, pp. 348–353.
[22] M. Orshansky, L. Milor, P. Chen, K. Keutzer, and C. Hu, "Impact of spatial intrachip gate length variability on the performance of high-speed digital circuits," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 21, no. 5, pp. 544–553, May 2002.

[23] S. Samaan, "The impact of device parameter variations on the frequency and performance of VLSI chips," in *Proc. IEEE/ACM ICCAD*, San Jose, CA, Nov. 2004, pp. 343–346.
[24] K. Bowman, S. Samaan, and N. Hakim, "Maximum clock frequency distribution model with practical VLSI design considerations," in *Int. Conf. Integrated Circuit Design Technology (ICICDT)*, Austin, TX, May 2004, pp. 183–191.
[25] *International Technology Roadmap for Semiconductors (ITRS)*. (2001). [Online]. Available. http://public.itrs.net
[26] N. C. Beaulieu, A. A. Abu-Dayya, and P. J. Mclane, "Estimating the distribution of a sum of independent log-normal random variables," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 2869–2873, Dec. 1995.
[27] R. Rao, A. Devgan, D. Blaauw, and D. Sylvester, "Parametric yield estimation considering leakage variability," in *Proc. IEEE/ACM DAC*, San Diego, CA, Jun. 2004, pp. 442–447.
[28] *Log Normal Distribution*. [Online]. Available: http://mathworld.wolfram.com/LogNormalDistribution.html
[29] *Maple 9.5*. [Online]. Available: http://www.maplesoft.com/products/maple/

Rajeev R. Rao received the B.S. degree in electrical and computer engineering from Rutgers University, New Brunswick, NJ, in 2002 and the M.S. degree in computer science and engineering from the University of Michigan, Ann Arbor, in 2004. He is currently working toward the Ph.D. degree at the University of Michigan.

In the summer of 2003, he was with IBM Austin Research Laboratory, Austin, TX, working as a Research Co-Op on leakage power analysis. His research interests include modeling and analysis of robust low-power very large scale integrated (VLSI) designs and variability-aware circuit approaches.

Anirudh Devgan (S'91–M'91–SM'00) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, in 1990, and the M.S. and Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 1991 and 1993, respectively.

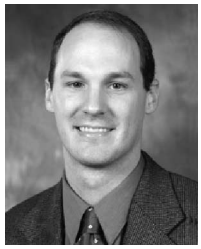
He has held various key technical and management positions at IBM Thomas J. Watson Research Center, IBM's Server Division, IBM's Microelectronics Division, and IBM Austin Research Laboratory. He is currently Vice President of Product Development at Magma Design Automation, Austin, TX. He has worked in the general area of computer-aided design of integrated circuits with emphasis on electrical analysis and simulation, physical design, low-power design, and design for manufacturability and variability. He has more than 60 publications, seven full-day tutorials, several invited presentations, and 20 U.S. patents issued or pending.

Dr. Devgan has been awarded the IEEE William J McCalla Award, IBM Corporate Award, IBM Outstanding Innovation Award, IBM Outstanding Research Accomplishment Award, IBM Microelectronics Division Excellence Award, and several other IBM awards. He has served on program committees of various international conferences including the Design Automation Conference (DAC), the International Conference on Computer-Aided Design (ICCAD), the International Conference on Computer Design (ICCD), the Asia South Pacific Design Automation Conference (ASP-DAC), VLSI Design, and the International Symposium on Quality Electronic Design (ISQED).

David Blaauw (M'94) 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 worked at IBM Corporation as a Development Staff Member, until August 1993. From 1993 until August 2001, he worked for Motorola, Inc., Austin, TX, were he was the Manager of the High Performance Design Technology Group. Since August 2001, he has been in the faculty of the University of Michigan as an Associate Professor. His research interests include very large scale integrated (VLSI) design and computer-aided design (CAD) with particular emphasis on circuit design and optimization for high-performance and low-power designs.

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



Dennis Sylvester (S'96–M'97–SM'04) received the Ph.D. degree in electrical engineering from the University of California, Berkeley, in 1999.

After working as a Senior R&D Engineer in the Advanced Technology Group of Synopsys, Mountain View, CA, he is now an Assistant Professor of electrical engineering and computer science at the University of Michigan, Ann Arbor. His research interests include low-power circuit design and design automation, design for manufacturability, and on-chip interconnect modeling.