

Modeling and Estimation of Full-Chip Leakage Current Considering Within-Die Correlation*

Khaled R. Heloue, Navid Azizi, Farid N. Najm
 Department of ECE, University of Toronto, Toronto, Ontario, Canada
 {khaled,nazizi,najm}@eecg.utoronto.ca

ABSTRACT

We present an efficient technique for finding the mean and variance of the full-chip leakage of a candidate design, while considering logic-structures and both die-to-die and within-die process variations, and taking into account the spatial correlation due to within-die variations. Our model uses a “random gate” concept to capture high-level characteristics of a candidate chip design, which are sufficient to determine its leakage. We show empirically that, for large gate count, the set of all chip designs that share the same high level characteristics have approximately the same leakage, with very small error. Therefore, our model can be used as either an *early* or a *late* estimator of leakage, with high accuracy. In its simplest form, we show that full-chip leakage estimation reduces to finding the area under a scaled version of the within-die channel length auto-correlation function, which can be done in constant time.

Categories and Subject Descriptors B.7.2 [Integrated Circuits]: Design Aids;

General Terms: Algorithms

Keywords: Statistical Analysis, Leakage Power.

1. INTRODUCTION

As a result of technology scaling, leakage current is becoming a major design challenge, affecting both circuit performance and power. Thus, estimating full-chip leakage becomes increasingly important. The leakage current of a circuit is not, however, simply the sum of the leakages of the devices in the circuit. Not only do logic-gate structures, such as stacking, affect the device leakage, but process variations make leakage estimation statistical in nature.

Full-chip leakage estimation is useful at different points in the design flow. Towards the end of the design flow (late mode estimation), leakage estimation can be used as a final sign-off tool, and requires a complete netlist with possibly a circuit placement. On the other hand, early estimation of leakage (early mode estimation) provides the full-chip leakage given limited information about the design, which is very useful to allow for design planning. While early work on leakage estimation concentrated on early mode estimators, these works [1, 2] either did not consider logic-gate structures and other transistor topologies, and/or did

*This work was supported in part by Intel Corp. and Altera Corp.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DAC 2007, June 4–8, 2007, San Diego, California, USA.
 Copyright 2007 ACM 978-1-59593-627-1/07/0006 ...\$5.00.

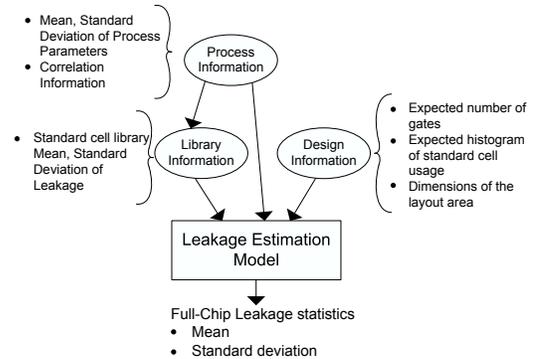


Figure 1: Leakage Estimation Model and the High-level characteristics required

not consider the effect of correlation between the variations on the total leakage. More recent work [3, 4] has taken into consideration both the effects of gate topologies and correlation. However these methods are late mode estimators of leakage that require minimally the circuit netlist and possibly a circuit placement to provide a leakage estimate, and they operate at the level of the netlist, so they can be expensive on large circuits, with a complexity of $\mathcal{O}(n^2)$ (some refinements are possible to reduce this cost, but with some loss of accuracy [3]). Given the need to budget for power constraints, there is a need for accurate early mode estimators that take into consideration both correlation and gate topologies. As for late mode estimators, more efficient techniques are required.

We present a new model and methodology for full-chip leakage estimation, in which certain high-level characteristics of a candidate chip design are used to determine its leakage statistics with high accuracy. For late mode estimation, these characteristics can be *extracted* from the netlist and/or placement. For early mode estimation, these characteristics can be simply specified as *expected* values based on previous design experience or on decisions made in the floorplanning stage. Our methodology uses a concept of a “random gate” to capture these characteristics and considers both correlations and gate topologies. We show that these high-level characteristics are sufficient to determine the leakage statistics of a design.

A block diagram of the system is shown in Fig. 1. Given information about (1) the process, (2) the standard cell library, and (3) certain high-level design characteristics, we predict the mean and standard deviation of full-chip leakage. The process information includes the mean and standard deviation of the underlying process variations, such as the variations in transistor length or threshold voltage, and information regarding the within-die spatial correlation. The standard cell library information includes the leakage characteristics of the cell library under process variations; this information can be obtained by pre-characterizing the cells in the library. Finally, some information on the candidate design is needed, including the (extracted or expected) cell usage histogram (*i.e.*, frequency of use distribution) for cells in the li-

brary, the (extracted or expected) number of cells in the design, and the dimensions of the layout area. With this, we determine the full-chip leakage statistics (mean and variance) for the design.

To carry out the estimation, we propose a model which is *generic*, in the sense that it is a *template* for all designs that share the same values for these high-level characteristics. We use probability theory as the vehicle to implement this template, so that all designs that share the same values of these high-level characteristics will be *members* or *instances* of this probabilistic template model. We introduce the concept of the Random Gate (RG) which allows us to capture the characteristics of a candidate design. This allows the leakage statistics to be obtained in $\mathcal{O}(n)$ time, where n is the number of cells in the design, but we then also show that, for large gate counts, the statistics of the full-chip leakage can be written in integral form, allowing for the computational complexity of our estimator to become $\mathcal{O}(1)$ time.¹ The key point, the thesis of this work, is that *large* designs that share the same high-level characteristics will have approximately the same leakage statistics and, by leveraging this property, our estimation engine provides accurate and efficient estimation, either early or late in the design flow.

2. MODELING

Variations normally have two components: a Die-to-Die (D2D) component, and a Within-Die (WID) component. The D2D component is a variation between different instances of the die and is shared by all devices on the same die. The WID component of variation, however, causes different devices on the same die to have different process parameters; the WID variations have some correlation across the die. D2D and WID variations are considered to be (statistically) independent, so that the total variance of a process parameter, such as transistor length, when both sources of variation are considered, can be written as $\sigma^2 = \sigma_{dd}^2 + \sigma_{wd}^2$ where σ_{dd}^2 is the variance of the D2D variation and σ_{wd}^2 is the variance of the WID variation. To model the WID spatial correlation between variations in transistor characteristics, we assume the existence of a spatial correlation function [5] that depends on the distance between the two transistors. Given the D2D and WID parameter variances, and the WID correlation, one can easily determine the total correlation between parameter variations (due to D2D and WID effects) by a simple normalization.

2.1 Cell Modeling

While the distribution of the underlying process parameters can be obtained from the foundry, the leakage distributions of each cell can not be immediately obtained. Since each cell has a different topology, with different transistor stacks, the leakage in each cell is affected differently by the underlying variations in the transistor length and threshold voltage. Furthermore the cell's inputs also affect the leakage distribution of each cell.

Leakage current is determined primarily by transistor, not interconnect, parameters. Of the many transistor parameters, the truly relevant ones are channel-length (L) and threshold voltage (V_t), due to the exponential dependence of leakage current on these two parameters. Threshold voltage variations are mainly due to two effects: random dopant fluctuations in the channel and the V_t roll-off effect whereby V_t varies in response to variations in L . For this work, when we refer to V_t variations, we specifically refer to the effect of random dopant fluctuations. We lump the effect of V_t roll-off on leakage into the L variations, because the two are directly related. This allows us to make the simple statement that V_t variations are purely random (independent) across the die [6], while L variations are not [3] (they include some within-die correlation). This approach is in line with the modern treatment of leakage in published work [2].

Since V_t variations are independent, while L variations are not, it follows immediately that, for full-chip leakage estimation, while V_t variations may be relevant for finding the *mean* of the total leakage, they are definitely not relevant for finding the *variance*

¹When used as a late mode estimator, there will be some additional cost to extract the cell usage histogram from the netlist, but that also can be constant-time, or linear-time in the worst case.

of the total leakage. The reason for this is simple: the variance of the sum of n independent random variables is $\sim n\sigma^2$, while the variance of the sum of n highly correlated random variables is $\sim n^2\sigma^2$. Thus, for large chips (large n), the variance of chip leakage due to V_t variations is negligible compared to that due to L variations. This too is in line with the modern published work on leakage [2]. Thus, for leakage *variance* estimation, we can focus on L alone. As for the effect of V_t variations on the *mean* leakage, that can be easily determined through a multiplicative term that depends on the variance of V_t , which is derived from the mean of the log-normal distribution, similar to [8]. As this is standard textbook material, it will not be covered here.

To model the distribution of the leakage of each cell, we use two methods which have different levels of computational complexity and accuracy. The first method uses a Monte-Carlo (MC) analysis to obtain the leakage statistics of each cell. While this technique needs extensive simulations, it does give us some confidence in the resulting distributions. The second method, an analytical approach, uses a limited sampling of the leakage of the cell, and then fits the leakage of the cell into a functional form, from which we compute the mean and variance of the distribution. These two methods are discussed below, and we then discuss correlation and circuit state dependency.

2.1.1 Monte-Carlo Technique

We use a commercial 90nm CMOS technology, along with its associated standard cell library of which we use 62 cells which include the Static Random Access Memory (SRAM) cell, various flip flops and a range of different logic cells. For each cell and input combination, we perform a MC analysis to determine the mean (μ) and standard deviation (σ) of the cell's leakage. The MC analysis is done assuming all the variations in the transistor channel length within the cell are completely correlated, which is reasonable in practice given that the transistors in each cell are very close together.

2.1.2 Analytical Technique

Rao et al. introduced [2] a mathematical model to express the leakage current, X , of a given cell as a function of channel length, L , to be $X = ae^{bL+cL^2}$ and showed that a fitted model can accurately model the leakage of different topologies including individual transistors and transistor stacks.

In our work, after we fit each cell's leakage into the same functional form, we use the triplet (a, b, c) to determine the mean and variance of the underlying leakage distribution exactly. The derivation, which is not shown due to space restrictions, results in:

$$\mu_{\mathbf{X}} = M_{\mathbf{Y}}(1) \quad (1)$$

$$\sigma_{\mathbf{X}}^2 = M_{\mathbf{Y}}(2) - \mu_{\mathbf{X}}^2 \quad (2)$$

where $M_{\mathbf{Y}}(t)$ is the moment-generating function of $\mathbf{Y} = \ln \mathbf{X}$ which can be shown to be:

$$M_{\mathbf{Y}}(t) = (1 - 2K_1 t)^{\frac{1}{2}} e^{\left[\frac{K_2^2 K_1 t}{1 - 2K_1 t} + K_3 t \right]} \quad (3)$$

by using the moment generating function of the "Non-Central Chi-square" distribution where K_1 , K_2 and K_3 are simple functions of the regression parameters (a, b, c) and the mean μ and standard deviation σ of the length, as follows:

$$K_1 = c\sigma^2 \quad K_2 = \frac{1}{\sigma} \left(\frac{b}{2c} + \mu \right) \quad (4)$$

$$K_3 = \ln a + b\mu + c\mu^2 - c \left(\frac{b}{2c} + \mu \right)^2 \quad (5)$$

To check the accuracy of the analytical model in determining the mean and standard deviation of cell's leakage, we compare the results obtained from the analytic model to the results obtained through MC analysis for all 62 cells with all input combinations. For the mean, the analytical method is quite close to the MC results; there is less than a 2% error for all gates, and the average absolute error is 0.44%. For the standard deviation, the average absolute error is 3.1%, and the maximum error is about 10%.

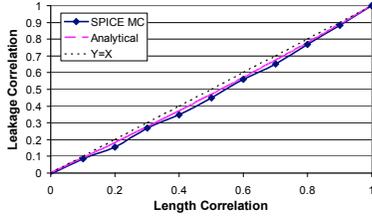


Figure 2: Correlation in length vs in leakage for different gates

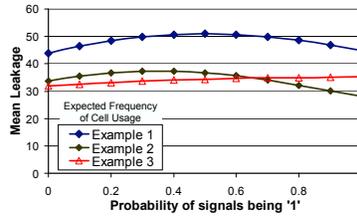


Figure 3: Effects of signal probability

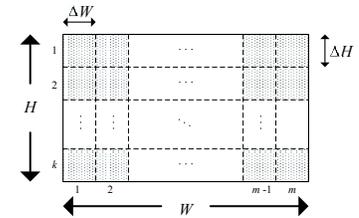


Figure 4: Abstract organization of die

The error in the mean and standard deviation is not a result of the mathematical derivation, but due to the leakage curve not being exactly mapped to the functional form ae^{bL+cL^2} . Thus, there is a trade-off between computational complexity and accuracy; if MC analysis is performed on all gates, then the distribution models for all gates will have high accuracy; on the other hand, using the functional form requires minimal simulation time.

2.1.3 Leakage Correlation

As mentioned, we assume the existence of a spatial correlation function which gives the correlation between *process parameters* as a function of the distance separating two locations, but which does not provide the correlation between the *leakages* of two cells at these locations. Using the regressed triplets, (a, b, c) , we have developed an analytical method that determines the leakage correlation between any pair of gates placed at two arbitrary locations on the die given the correlation in their channel lengths. In other words, we have determined a *mapping* $\rho_{m,n}(l_i, l_j) = f_{m,n}(\rho_L(l_i, l_j))$ where $\rho_L(l_i, l_j)$ is the channel length correlation between two locations l_i and l_j , $f_{m,n}(\cdot)$ is the derived mapping for gates m and n and $\rho_{m,n}(l_i, l_j)$ is the leakage correlation for gates m and n placed at locations l_i and l_j respectively.

The details of this mapping are not shown for lack of space, but Fig. 2 shows the results of the leakage correlation given a length correlation for both the MC analysis and the analytical technique for a single pair of gates; note that the analytical technique shows a good match to the MC results. Also the leakage correlation is near the $y = x$ line, at which leakage correlation equals channel length correlation. We have performed the analysis for all pairs of gates, and shown that the analytical mapping provides accurate results in all cases. The set of mappings $f_{m,n}(\cdot)$ for different gates are slightly different but they all closely follow the $y = x$ line. We will use this observation that the leakage correlation is close to the length correlation in the case where MC analysis is used to obtain the cell leakage statistics since we do not have the (a, b, c) triplet to obtain the leakage correlation exactly.

2.1.4 Input Combinations

The signal probability (probability that a logic signal is 1) certainly has an effect on leakage. This effect is quite strong for single logic gates, causing a spread of 10X in some cases. However, for large circuits, the impact of signal probability is significantly diminished due to averaging of their effects (law of large numbers). To study this effect, we have swept the signal probabilities from 0 to 1 and have found, as shown in Fig. 3, that the effect on large circuit leakage is not pronounced and is also dependent on the frequency by which various cells are employed in the design. The figure shows the leakage mean, and similar behavior has been found for the leakage variance. For a practical solution approach, one has the option of simply setting the signal probabilities at some ball-park mid-level value, such as 0.5. A better approach, which we employ, is to first characterize every cell for all its input states; then, based on this pre-characterized data, and for the given frequency of use distribution for cells, find the signal probability setting which maximizes the mean leakage, effectively finding the maximum of a plot such as Fig. 3. Empirically, we find that this setting turns to be very good for finding the maximum leakage mean for the candidate design, as well as its maximum leakage variance. This approach gives a conservative estimate, in the face of uncertainty about eventual signal probabilities.

2.2 Full-Chip model

What determines the leakage of a large circuit? We will demonstrate empirically that certain high-level characteristics of a candidate design are sufficient to determine its leakage. *In a library-based standard-cell design environment, these characteristics are: 1) the cell library (characterized for leakage), 2) the (actual or expected) frequency of usage for cells in the library, 3) the (actual or expected) number of cells in the design, and 4) the dimensions of the layout area.* In order to carry out the leakage estimation, we propose a model for the candidate chip design which is *generic*, in the sense that it is a *template* for all designs that share the same values for these high-level characteristics. We use probability theory as the vehicle to construct this template, so that all designs that share the same values of these high-level characteristics will be *members* or *instances* of this probabilistic template model. After developing our leakage predictor based on this model, we will then show that the leakages of all instances of specific designs which are members of this model converge towards the predicted leakage value as the circuit size increases; Fig. 6 offers a “sneak preview” of this convergence.

2.2.1 Model Definition and Suitability

Formally, our full-chip model is a rectangular array of a number (n) of identical *sites*, as shown in Fig. 4, where every site is occupied by a probabilistic abstraction which we call a *random gate* (RG), and such that the dimensions of the array are equal to the dimensions of the layout area of the candidate design, and that the number of sites n is equal to the number of cells in the design. But what is a RG? Simply put, a RG is similar to a Random Variable (RV); however, unlike a RV which assumes real numbers as *outcomes* or *instances*, the instances of a RG are gates from the standard-cell library, with probabilities identical to those in the frequency of use distribution. In other words, the RG discrete probability distribution is identical to the frequency of cell usage of the design.

This full-chip array model is a suitable probabilistic representation of all designs having the high-level characteristics highlighted earlier. On one hand, its dimensions and gate count match the dimensions of the layout and the number of cells in the candidate design. On the other hand, the frequency of cell usage of the design is also matched by the way the RG discrete probability distribution is defined. Hence, if an *instance* of the full-chip model is defined to be n RG instances at every site in the array, then the frequency of cell usage for that full-chip model *instance* will be identical to the frequency of cell usage of the candidate design, for large n . Therefore, the full-chip model is a probabilistic representation of a set of designs with the same high-level characteristics, and those designs are in fact *instances* of our model. Using this fact, we will use the full-chip model to estimate the leakage of the candidate design.

One possible reaction to this proposal is that all sites in the full-chip model are of identical size while obviously cells in the library are of different sizes. Another comment is that the array seems to leave no room for interconnect routing. Both these issues do not present a problem. In fact, the size of a site is really the size of the layout area, divided by the number of cells, thus it is the average size of a cell and the interconnect that may be associated with it. Thus, all that is captured by the notion of a RG site is the idea that the leakage due to one cell would on average be spread out or “allocated” to the layout area of a single site.

2.2.2 Leakage Statistics of a Random Gate

As stated earlier, the RG is simply a gate picked at random from the library, according to a discrete probability distribution which is identical to the frequency of gate usage. In order to perform full-chip leakage estimation based on our model, we need to construct and mathematically define the leakage statistics of the RG.

Let \mathbf{I} be an RV that takes as values the *type* of a gate picked from the library at random to be used in the design. This means that $\mathbf{I} \in \{1, 2, \dots, p\}$, where p is the total number of gates in the library, and that the distribution of \mathbf{I} is identical to the frequency of gate usage. Let α_i be the frequency of usage of gate i . Then:

$$\mathcal{P}\{\mathbf{I} = i\} = \alpha_i \quad \forall i = 1, 2, \dots, p \quad \text{and} \quad \sum_{i=1}^p \alpha_i = 1 \quad (6)$$

Let $\mathbf{X}_{\mathbf{I}}$ be an RV that represents the leakage of a gate picked according to the distribution of \mathbf{I} . Then by definition, $\mathbf{X}_{\mathbf{I}}$ is the leakage of the RG. Consequently, $\mathbf{X}_{\mathbf{I}}$ is defined on two probability spaces; the space of \mathbf{X} due to channel length variations, and the space of \mathbf{I} due to the choice of gate type. Note that for an arbitrary realization of say $\mathbf{I} = i$, $\mathbf{X}_{\mathbf{I}}$ will be equal to \mathbf{X}_i , that is the RV that represents the leakage of gate of type i . Recall that the statistics of \mathbf{X}_i , *i.e.*, its mean μ_i and standard deviation σ_i , have already been determined during pre-characterization for all gates i in the library. We can determine the mean leakage $\mu_{\mathbf{X}_{\mathbf{I}}}$ of the RG as follows:

$$\mu_{\mathbf{X}_{\mathbf{I}}} = E[\mathbf{X}_{\mathbf{I}}] = E_I[E_X[\mathbf{X}_{\mathbf{I}} | \mathbf{I}=i]] = E_I[E_X[\mathbf{X}_i]] = \sum_{i=1}^p \alpha_i \mu_i \quad (7)$$

where $E_X[\cdot]$ and $E_I[\cdot]$ are the expected values over the spaces of \mathbf{X} and \mathbf{I} , respectively. To determine the variance $\sigma_{\mathbf{X}_{\mathbf{I}}}^2$ of $\mathbf{X}_{\mathbf{I}}$, we start by determining its second moment $E[\mathbf{X}_{\mathbf{I}}^2]$ as:

$$E[\mathbf{X}_{\mathbf{I}}^2] = E_I[E_X[\mathbf{X}_{\mathbf{I}}^2 | \mathbf{I}=i]] = E_I[E_X[\mathbf{X}_i^2]] = \sum_{i=1}^p \alpha_i (\sigma_i^2 + \mu_i^2) \quad (8)$$

Given the second moment and the mean, the variance can be trivially determined as $E[\mathbf{X}_{\mathbf{I}}^2] - \mu_{\mathbf{X}_{\mathbf{I}}}^2$.

2.2.3 Random Gate Leakage Correlation

In addition to the RG leakage statistics defined in the previous section, we need to construct and define the RG leakage correlation.

Recall that $\mathbf{X}_{\mathbf{I}}$ is defined as the leakage of a random gate picked from the library according to the distribution of \mathbf{I} , and placed at some location on the die. Let $\mathbf{X}_{\mathbf{I}}(l_i)$ and $\mathbf{X}_{\mathbf{I}}(l_j)$ be the leakages of the two RGs at two arbitrary locations l_i and l_j . It is important to understand that $\mathbf{X}_{\mathbf{I}}(l_i)$ and $\mathbf{X}_{\mathbf{I}}(l_j)$ are identically distributed, and any correlation among these RVs is only due to the correlation over the space of process variations and not over the space of gate selection.

Let $C_{\mathbf{X}_{\mathbf{I}}}(l_i, l_j)$ be the covariance of $\mathbf{X}_{\mathbf{I}}(l_i)$ and $\mathbf{X}_{\mathbf{I}}(l_j)$, which is defined as $C_{\mathbf{X}_{\mathbf{I}}}(l_i, l_j) = E[\mathbf{X}_{\mathbf{I}}(l_i)\mathbf{X}_{\mathbf{I}}(l_j)] - \mu_{\mathbf{X}_{\mathbf{I}}}^2$. It can be shown, using conditional expectation, that this covariance is given by:

$$C_{\mathbf{X}_{\mathbf{I}}}(l_i, l_j) = \sum_{m=1}^p \sum_{n=1}^p \alpha_m \alpha_n C_{m,n}(l_i, l_j) \quad (9)$$

where $C_{m,n}(l_i, l_j)$ is the covariance of the leakage of two gates of types m and n , when placed at locations l_i and l_j , respectively, *i.e.*, $\mathbf{X}_m(l_i)$ and $\mathbf{X}_n(l_j)$. Note that the covariance of the leakage of the random gate $\mathbf{X}_{\mathbf{I}}$ is the expected value over \mathbf{I} of the covariances of all pairs of gate types. This result is somewhat intuitive since the random gate is an abstraction that embodies all gates in the library. Starting from (9), we can normalize $C_{m,n}(l_i, l_j)$ by the standard deviations of gates m and n to get their leakage correlation $\rho_{m,n}$. Then, we use the analytical mapping $f_{m,n}(\cdot)$ from Section 2.1.3 to relate the leakage correlation $\rho_{m,n}$ to channel length correlation ρ_L , as follows:

$$\begin{aligned} C_{\mathbf{X}_{\mathbf{I}}}(l_i, l_j) &= \sum_{m=1}^p \sum_{n=1}^p \alpha_m \alpha_n [\rho_{m,n}(l_i, l_j) \sigma_m \sigma_n] \\ &= \sum_{m=1}^p \sum_{n=1}^p \alpha_m \alpha_n \sigma_m \sigma_n f_{m,n}(\rho_L(l_i, l_j)) \end{aligned} \quad (10)$$

Let $F(\rho_L(l_i, l_j))$ be equal to the final expression in (10) above, and notice that this equation assumes that l_i and l_j are different. When they are the same, $C_{\mathbf{X}_{\mathbf{I}}}(l_i, l_j)$ is just the variance $\sigma_{\mathbf{X}_{\mathbf{I}}}^2$. Thus:

$$C_{\mathbf{X}_{\mathbf{I}}}(l_i, l_j) = \begin{cases} F(\rho_L(l_i, l_j)) & \text{for } l_i \neq l_j \\ \sigma_{\mathbf{X}_{\mathbf{I}}}^2 & \text{for } l_i = l_j \end{cases} \quad (11)$$

By enforcing this correlation structure on our RG array, we ensure that instances of this array have the same correlation structure as the candidate design.

3. FULL-CHIP LEAKAGE ESTIMATION

For a specific placed design, based on a pre-characterized cell library, one can determine the full-chip leakage statistics using techniques from standard probability theory [7] for finding the sum of a number of correlated RVs (each RV corresponds to the leakage of one cell instance). This would be an $\mathcal{O}(n^2)$ approach, which can be expensive for large circuits (some refinements are possible to reduce this cost, but with some loss of accuracy [3]). Throughout this paper, we will refer to the leakage obtained from such an $\mathcal{O}(n^2)$ approach as the *true leakage* of a given design.

Apart from the issue of computational cost, such an approach is available only later in the design flow once a netlist and placement are available; it is useful only as a final check, and not as a prelude to corrective action. In this section, we will first show how we can determine the full-chip leakage statistics in linear time, $\mathcal{O}(n)$, and then show how this can be improved to obtain the statistics in constant time, $\mathcal{O}(1)$. Importantly, we will also show that, for large gate counts, the statistics of any specific design that shares the same high-level characteristics under consideration converge to the values predicted by our model.

3.1 Linear-time method

Let \mathbf{I}_T be an RV that represents the leakage of our full-chip model, *i.e.*, of the array of n RGs. This means that:

$$\mathbf{I}_T = \sum_{i=1}^n \mathbf{X}_{\mathbf{I}}(l_i) \quad (12)$$

where l_i is the location of the i^{th} random gate. We are interested in determining the statistics of \mathbf{I}_T , namely its mean $\mu_{\mathbf{I}_T}$ and variance $\sigma_{\mathbf{I}_T}^2$. The mean of \mathbf{I}_T is equal to:

$$\mu_{\mathbf{I}_T} = E[\mathbf{I}_T] = \sum_{i=1}^n E[\mathbf{X}_{\mathbf{I}}(l_i)] = \sum_{i=1}^n E[\mathbf{X}_{\mathbf{I}}] = n \mu_{\mathbf{X}_{\mathbf{I}}} \quad (13)$$

The variance of \mathbf{I}_T can be easily determined using a result from probability theory that the variance of a sum of correlated RVs is equal to the sum of pairwise covariances [7]. In other words:

$$\sigma_{\mathbf{I}_T}^2 = \sum_{a=1}^n \sum_{b=1}^n C_{\mathbf{X}_{\mathbf{I}}}(l_a, l_b) \quad (14)$$

Note that the above double summation accounts also for the cases where $l_a = l_b$, for which the covariance is essentially the variance. Using the fact that any covariance can be written in terms of the correlation, $C_{\mathbf{X}_{\mathbf{I}}}(l_a, l_b) = \rho_{\mathbf{X}_{\mathbf{I}}}(l_a, l_b) \sigma_{\mathbf{X}_{\mathbf{I}}}^2$, we can write the total leakage variance in its final form:

$$\sigma_{\mathbf{I}_T}^2 = \sigma_{\mathbf{X}_{\mathbf{I}}}^2 \sum_{a=1}^n \sum_{b=1}^n \rho_{\mathbf{X}_{\mathbf{I}}}(l_a, l_b) \quad (15)$$

where the variance of the full-chip leakage is a function of the variance of the random gate and the extent of leakage correlation across the chip.

At this point, we have determined the mean of the total leakage (in constant time), and have shown that the computation of the variance of the total leakage requires a double summation over the number of gates on the chip. This $\mathcal{O}(n^2)$ complexity is not practically acceptable, especially knowing that n can be extremely large, on the order of millions. By taking into account the shape of the die and the sole dependence of the leakage correlation on the distance between different locations, we are able to cut down the complexity of computing the total leakage variance to $\mathcal{O}(n)$, as follows.

Let the RG array consist of k rows and m columns, where the total number of gates, n , is equal to the product $k \times m$, as shown in Fig. 4. Each location or "site" on the grid can be represented by a pair (r, s) where r is the horizontal index taking values $r = 1, \dots, m$ and s is the vertical index taking values $s = 1, \dots, k$. Also, assume that the height H and width W of

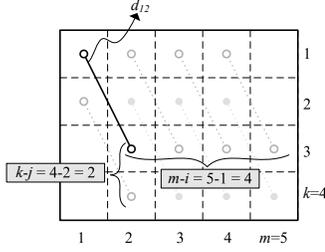


Figure 5: Number of occurrences of a certain distance vector

Table 1: % Error in full-chip Standard Deviation for ISCAS85 circuits compared to the RG estimates

c499	c1355	c432	c1908	c880	c2670	c5315	c7552	c6288
1.04%	0.41%	1.14%	0.36%	0.74%	0.52%	0.23%	0.34%	1.38%

the array are known. Let ΔH and ΔW be the height and width of the site where every gate will be placed.

Given the above parameters, the centre to centre distance d_{ij} between any two sites (r_1, s_1) and (r_2, s_2) can be easily determined to be $d_{ij} = \sqrt{(i \cdot \Delta W)^2 + (j \cdot \Delta H)^2}$ where i is defined as the algebraic difference in horizontal indices, *i.e.*, $(r_2 - r_1)$, and j is defined as the algebraic difference in vertical indices, *i.e.*, $(s_2 - s_1)$. Note that $i = 0, \pm 1, \dots, \pm(m-1)$ and $j = 0, \pm 1, \dots, \pm(k-1)$.

Now recall the total leakage variance defined in (15) where the double summation covers all possible pairs of locations, and each location is a site on the grid defined by two indices. Since the correlation depends only on the distance d_{ij} between the pairs of locations, we can simplify the above expression greatly by performing the sum over the different distances rather than the pairs of locations. To do that, however, we need to determine the number of times each distance d_{ij} occurs. This is relatively easy for a rectangular $k \times m$ grid, as can be seen in Fig. 5, where the number of times a distance d_{ij} occurs along the width of the die is $m - |i|$ and along the height of the die is $k - |j|$. Using these two values, the number of occurrences n_{ij} of d_{ij} can be determined to be the following:

$$n_{ij} = (m - |i|) \cdot (k - |j|) \quad (16)$$

Since the leakage correlation between any two given locations depends only on the distance between these locations, we will explicitly highlight this fact, $\rho_{\mathbf{X}_I}(l_a, l_b) = \rho_{\mathbf{X}_I}(d_{ij})$ where i and j in the above equation are the algebraic differences in the horizontal and vertical indices of l_a and l_b .

Starting from (15), we will transform the quadratic summation that runs over all pairs of locations, into a summation that runs over the set of possible distances induced by the rectangular shape of the grid. This set will be covered if all the algebraic differences i and j are covered. After accounting for the number of times each algebraic difference occurs, n_{ij} , we get the following expression for the total leakage variance:

$$\sigma_{\mathbf{I}_T}^2 = \sigma_{\mathbf{X}_I}^2 \sum_{i=-m}^m \sum_{j=-k}^k (m - |i|) \cdot (k - |j|) \rho_{\mathbf{X}_I}(d_{ij}) \quad (17)$$

where the double summation runs at most $\mathcal{O}(k \times m) = \mathcal{O}(n)$ times. This summation is linear in circuit size. Note that the expression in (17) is an exact transformation of (15) without any approximations.

3.1.1 Validation

Two types of validation tests were run, by first considering randomly generated circuits, as a way to make conclusions about the set of all circuits of a given size, and then by considering specific benchmark circuits.

In the first set of experiments, a large number of circuits were randomly generated so as to match a frequency of cell usage that was specified *a priori*. The circuits were then placed and routed, and their true leakage statistics (mean and variance) were found.

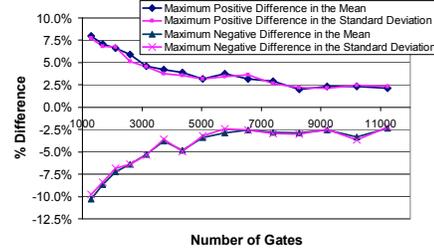


Figure 6: Errors in the estimation of mean and standard deviation of full-chip leakage

Fig. 6 shows the maximum positive and negative difference between the means and standard deviations of the leakages of these circuits compared to the estimates provided by our model. It can be seen that as the number of gates in the circuits increases, the difference approaches zero; at a circuit size of 11,236 gates, the maximum difference is 2.2%. This small amount of error indicates that the set of all chip designs that share the same high level characteristics have approximately the same full chip leakage statistics and thus these high-level characteristics are sufficient to determine chip leakage. This first set of experiments serves to justify the statement that this approach is useful as an *early estimator* of full-chip leakage.

In the second set of experiments, we show how the model can be used as a *late estimator* of leakage for real (placed and routed) circuits. In this test, we have *extracted* the relevant high-level characteristics from each ISCAS85 circuit, namely the number of gates used, the histogram of cells used, and the dimensions of the layout; then with these values, we have used our model to estimate the leakage statistics of every circuit. Table 1 lists the errors in the full-chip leakage standard deviation, for all ISCAS85 circuits, between our model and the true leakage of these circuits. The errors are very small (notice, however, that these do not include any cell leakage modeling errors, which were discussed earlier in section 2.1). We do not show the errors in the mean leakage because they are truly negligible.

3.1.2 Simplified Correlation Assumption

In Section 2.1 we noted that the cell leakage statistics (*i.e.*, the mean and standard deviation of leakage) can be obtained in two ways; either (1) a MC analysis would be done or (2) the cell's leakage would be fitted into a functional form to get three fitting parameters (a, b, c) . Using these parameters, the leakage mean and standard deviation were analytically obtained. The fitted parameters also allowed us to determine the leakage correlation between any pair of gates, $\rho_{m,n}$, given the channel length correlation ρ_L . Using the mapping, $f_{m,n}(\cdot)$, the RG leakage correlation was determined in (10).

If we, however, choose to obtain the leakage statistics of each cell through MC analysis, we would not be able to use $f_{m,n}(\cdot)$ to determine the leakage correlation between pairs of cells because the correlation mapping depends on the fitting parameters which are not available in MC mode. Without this mapping, the RG leakage correlation cannot be determined. The solution to this problem lies in Fig. 2, where we have noted that the leakage correlation of any pair of cells is approximately equal to the correlation in the channel length of these cells. In other words, $\rho_{m,n} \approx \rho_L, \forall m, n$. With this simplified correlation assumption, (10) can be used to determine the RG leakage correlation.

To determine the amount of error introduced by this assumption, we have compared the difference between the standard deviation when assuming $\rho_{m,n} = \rho_L$ compared to the analytical approach, *i.e.*, when using the true $f_{m,n}(\cdot)$ mapping. Regardless of whether we assume solely WID variations or have both WID and D2D variations, the percentage error is below 2.8%.

3.2 Constant-time method

In this section, we show how, for large values of n , we can approximate the linear summation in (17) by an integral to obtain the statistics of full-chip leakage in constant time.

3.2.1 2D Integration in Rectangular Coordinates

Starting from (17), let $x_i = i \cdot \Delta W$ and $y_j = j \cdot \Delta H$, and by multiplying out ΔW and ΔH we obtain:

$$\sigma_{I_T}^2 = \frac{\sigma_{\mathbf{X}_I}^2}{\Delta W \cdot \Delta H} \sum_{i=-m}^m \sum_{j=-k}^k (W - |x_i|) \cdot (H - |y_j|) \rho_{\mathbf{X}_I}(d_{ij}) \quad (18)$$

where $W = m \cdot \Delta W$, $H = k \cdot \Delta H$, and $d_{ij} = \sqrt{x_i^2 + y_j^2}$. By using a double integral to approximate the double summation over discrete values, we obtain:

$$\sigma_{I_T}^2 \approx \frac{\sigma_{\mathbf{X}_I}^2}{(\Delta W \cdot \Delta H)^2} \int_{x=-W}^W \int_{y=-H}^H (W - |x|) \cdot (H - |y|) \rho_{\mathbf{X}_I}(\sqrt{x^2 + y^2}) dy dx \quad (19)$$

Let the area of a RG site be $\mathcal{A}_{\text{site}} = \Delta W \Delta H$ and the area of the die be $\mathcal{A} = n \mathcal{A}_{\text{site}}$. Note that the function being integrated is even, so that we can write:

$$\sigma_{I_T}^2 \approx 4 \cdot \sigma_{\mathbf{X}_I}^2 \frac{n^2}{\mathcal{A}^2} \int_0^W \int_0^H (W - x) \cdot (H - y) \rho_{\mathbf{X}_I}(\sqrt{x^2 + y^2}) dy dx \quad (20)$$

The expression in (20) approximates the full-chip leakage variance for large values of n . Since the number of gates on the chip is typically in the order of millions, the approximation is valid in most cases. What is interesting about this expression is that it only requires the computation of an integral, which can be performed in constant-time using a good numerical integration routine; the leakage variance computation does not depend on the number of gates n , it is $\mathcal{O}(1)$.

3.2.2 1D Integration in Polar Coordinates

To make our computation even more efficient, under certain conditions we can transform the double integral in (20) into a single integral in polar coordinates. First we write an exact mapping of (20) in double-integral form using polar coordinates:

$$\sigma_{I_T}^2 \approx 4 \cdot \sigma_{\mathbf{X}_I}^2 \frac{n^2}{\mathcal{A}^2} \int_0^{\pi/2} \int_0^{D(\theta)} (W - r \cdot \cos \theta) \cdot (H - r \cdot \sin \theta) \rho_{\mathbf{X}_I}(r) r dr d\theta \quad (21)$$

where $D(\theta)$ is the distance from the origin to the boundary of the rectangular integration domain, which is less than the largest distance on the array. If the distance at which the WID correlation function reaches 0 is less than the minimum of the height or width of the array, then the double integral in (21) can be written as a single integral. To derive this single integral, let us for the moment assume that there are no D2D variations and that $\rho_{\mathbf{X}_I}$ becomes zero at a distance D_{max} . If D_{max} is less than $\min(W, H)$ then (20) can be written as:

$$\sigma_{I_T}^2 \approx 4 \cdot \sigma_{\mathbf{X}_I}^2 \frac{n^2}{\mathcal{A}^2} \int_0^{D_{max}} \int_0^{\pi/2} (W - r \cdot \cos \theta) \cdot (H - r \cdot \sin \theta) \rho_{\mathbf{X}_I}(r) r d\theta dr \quad (22)$$

Since the correlation function does not depend on θ , we can further simplify the above expression by separating the integrals:

$$\sigma_{I_T}^2 \approx 4 \cdot \sigma_{\mathbf{X}_I}^2 \frac{n^2}{\mathcal{A}^2} \int_0^{D_{max}} \rho_{\mathbf{X}_I}(r) r \left[\int_0^{\pi/2} (W - r \cdot \cos \theta) \cdot (H - r \cdot \sin \theta) d\theta \right] dr \quad (23)$$

The expression in the brackets can be analytically integrated and results in the following expression:

$$g(r) = 0.5r^2 - (W + H)r + \frac{\pi}{2}WH \quad (24)$$

which leads to the final expression for full-chip leakage variance:

$$\sigma_{I_T}^2 \approx 4 \cdot \sigma_{\mathbf{X}_I}^2 \frac{n^2}{\mathcal{A}^2} \int_0^{D_{max}} \rho_{\mathbf{X}_I}(r) \cdot r \cdot g(r) dr \quad (25)$$

When also considering D2D variations, recall from Section 2 that the correlation never reaches zero, and thus the single integral technique does not immediately apply. However, if we divide up the correlation function $\rho_{\mathbf{X}_I}(r)$ into a constant portion, ρ_C , and a portion that does go to 0 at D_{max} , $\rho'_{\mathbf{X}_I}(r) = \rho_{\mathbf{X}_I}(r) - \rho_C$, then the single integral can be written as:

$$\sigma_{I_T}^2 \approx \left[4 \cdot \sigma_{\mathbf{X}_I}^2 \frac{n^2}{\mathcal{A}^2} \int_0^{D_{max}} \rho'_{\mathbf{X}_I}(r) \cdot r \cdot g(r) dr \right] + \sigma_{\mathbf{X}_I}^2 n^2 \cdot \rho_C \quad (26)$$

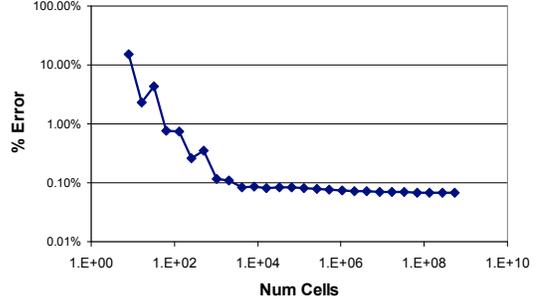


Figure 7: % Error between numerical integration and linear time algorithm

3.2.3 Validation

The value of the standard deviation of the full-chip leakage obtained from the numerical integration (20) was compared to the value obtained from the $\mathcal{O}(n)$ approach presented in Section 3.1.

As can be seen in Fig. 7, for circuits that have more than ten thousand gates there is less than 0.01% error between the numerical integration and that of the linear-time algorithm. For circuits with a small number of gates (<100) the % error is more than 1%; this is due to the granularity of the gates being a significant proportion of the total area of the design causing the integral to be less accurate than the true sum. For larger designs, the area of the logic gates compared to the area of the design approaches zero, allowing the numerical integration to provide good results, with less than 0.1% error.

Given that the $\mathcal{O}(n)$ time algorithm takes less than one second for circuits with less than 1000 gates, one can use the $\mathcal{O}(n)$ time algorithm in those cases, and use the numerical integration for circuits with a much larger number of gates.

4. CONCLUSION

We presented a probabilistic full-chip model that can be used to estimate, in constant-time, the leakage statistics of candidate designs either at an early or a late stage, while considering within-die correlations. We proposed and verified that certain high-level characteristics of a candidate chip design are sufficient to determine its leakage. These high-level characteristics, shown in Fig. 1, include information about the process, the standard-cell library, and the design in question. We showed that, for large gate count, the set of all chip designs that share the same high level characteristics have approximately the *same* full-chip leakage statistics, with very small error. We capture this set by a full-chip model based on Random Gates (RGs).

5. REFERENCES

- [1] S. Narendra, V. De, D. Antoniadis, and A. Chandrakasan. Full-chip sub-threshold leakage power prediction model of sub-0.18 μm CMOS. *ISLPED*, 2002.
- [2] R. Rao, A. Srivastava, D. Blaauw, and D. Sylvester. Statistical analysis of subthreshold leakage current for VLSI circuits. *TVLSI*, 12(2):131–139, February 2004.
- [3] H. Chang and S. S. Sapatnekar. Full-chip analysis of leakage power under process variations, including spatial correlations. *DAC*, 2005.
- [4] A. Agarwal, K. Kang, and K. Roy. Accurate estimation and modeling of total chip leakage considering inter- & intra-die process variations. *ICCAD*, 2005.
- [5] J. Xiong, V. Zolotov, and L. He. Robust extraction of spatial correlation. *ISPD*, 2006.
- [6] A. Keshavarzi, et al. Measurements and modeling of intrinsic fluctuations in MOSFET threshold voltage. *ISLPED*, 2005.
- [7] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, NY, 2nd edition, 1984.
- [8] D. Helms, et al. Analysis and modeling of subthreshold leakage of RT-components under PTV and state variation. *ISLPED*, 2006.