Efficient Additive Statistical Leakage Estimation

Lerong Cheng, Puneet Gupta, and Lei He

Abstract—Nominal power estimation is quick but gives minimal information. Statistical power analysis can provide information on yield, chip robustness, etc., but current methods are unnecessarily slow and complex. This is primarily because existing leakage-power models, which model leakage power as lognormal distribution and calculate chip leakage power based on Wilkinson’s approach, are not directly additive. Hence, for each incremental change of the circuit, the covariances between each pair of circuit elements need to be recalculated, which is inefficient. In this paper, we proposed a simple additive polynomial leakage-variation model. With additivity, we can calculate chip leakage power and leakage power after incremental change very efficiently. Experimental results show that our method is five times faster than the existing Wilkinson’s approach while having no accuracy loss in mean estimation and about 1% accuracy loss in standard-deviation and 99%-percentile-point estimations.

Index Terms—Leakage power, process variation, yield modeling.

I. INTRODUCTION

Nominal power analysis is and has been the standard way to estimate chip power. Although minimal information is obtained through this method (nominal chip power), it is both fast and incremental (requiring one simple addition per gate). On the other hand, statistical power analysis can provide the designer with a distribution of possible chip power consumption. This is very useful in estimating the parametric yield of a chip, thus directly translating into an estimation of profitability. Although this information is of great interest to chip manufacturers, there is currently no accurate statistical power modeling technique that is both fast and incremental. This makes statistical power analysis and optimization for large state-of-the-art chips a tedious task.

There have been many published approaches that cut dependence on full-circuit simulation altogether and are able to get decent results fairly fast [2]–[5]. Most of these works have been performed in regard to leakage power because it has the largest variance (it can vary 20 times while delay only changes 30%). These approaches still require complicated manipulation of their input parameters. Due to complex and interdependent parameter manipulation, these approaches are not incremental. Thus, even an $O(N)$ nonincremental algorithm would still be costly. Statistical power analysis through addition of distributions is another topic of research; however, these approaches all try to deal with the same snag. Leakage power is best characterized by a lognormal distribution; however, there is no closed-form solution to the addition of lognormals. Many of these approaches rely on Wilkinson’s approach to lognormal addition. Although these approaches find accurate solutions relatively quickly for smaller chips, any model that is primarily dependent on Wilkinson’s approach ($O(N^2)$) will not be scalable to large chips. An excellent example of this type of approach can be found in [3]. Reference [4] seeks to extract low-ranking quadratic models for each gate and hierarchically combines them into a low-ranking quadratic model for the whole chip. Such a method is fast when computing full-chip leakage power; however, it cannot handle incremental changes efficiently.

There are some approaches that are incremental and linear [6]. Reference [6] tries to model leakage variation as a set of orthogonal polynomials. Such a method is incremental and time efficient. However, it requires a different set of orthogonal polynomials for different distributions of variation sources, which makes it not flexible for different types of variation sources.

As discussed previously, Wilkinson’s approach, which most of the existing works are based on, is not directly additive. In order to calculate chip leakage power, the covariances of each pair of circuit elements need to be calculated. Moreover, during leakage-power optimization, some of the circuit elements of a chip may be incrementally changed. For Wilkinson’s approach, the covariances between each pair of the changed circuit elements and the remaining circuit elements need to be recalculated, which is very inefficient. However, if we have a leakage-power model that is directly additive, we may greatly improve the efficiency of chip-leakage-power calculation and optimization. With additivity, we only need simple linear operations for both chip-leakage-power and incremental calculations.

To achieve the additivity goal, we express leakage power as a polynomial instead of an exponential function of variation sources so that it is additive. It is shown that, in order to recalculate the chip leakage variation after changing some circuit elements, the computational complexity of our approach does not depend on the number of circuit-element types (standard-cell-library size), while the complexity of Wilkinson’s approach is linear to the circuit-element types.

The rest of this paper is organized as follows. Section II introduces the traditional Wilkinson’s approach. Section III presents our additive leakage-variation model. Section IV further extends our model with the existence of within-die variation. Section V shows some experimental results, and finally, Section VI concludes this paper.

II. WILKINSON’S APPROACH

Typically, leakage power is modeled as an exponential function of variation sources, $P = P_0 \cdot e^{\sum c_i X_i}$, where $P_0$ is the nominal leakage value, $X = (X_1, X_2, \ldots X_n)$ is the vector of variation sources, (e.g., effective channel length, threshold voltage, oxide thickness, etc.), and $c_i$’s are coefficients that can be obtained from curve fitting through SPICE simulation or measurement. In most cases, variation sources are assumed to be with Gaussian distribution. Therefore, cell leakage power is with lognormal distribution. With the single-cell leakage-power-variation model, full-chip leakage power is calculated as the sum of the leakage powers of all cells, $P_{\text{chip}} = \sum_{i \in T} N_i P_i$, where $T$ is the set of all cell types in the chip, $N_i$ is the number of the $i$th cell, and $P_i$ is the leakage power of the $i$th cell. As discussed earlier, $P_i$ is modeled as a lognormal random variable. Then, $P_{\text{chip}}$ is the sum of lognormal random variables. However, there is no closed-form expression to calculate the distribution of the sum of lognormal distributions. The most popular method is to approximate the sum of

Manuscript received September 25, 2008; revised February 25, 2009 and August 6, 2009. Current version published October 21, 2009. This work was supported in part by Integrated Modeling Process and Computation for Technology (IMPACT). This paper was recommended by Associate Editor F. N. Najm.

The authors are with the University of California at Los Angeles, Los Angeles, CA 90095 USA (e-mail: lerong@ee.ucla.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCAD.2009.2030433

0278-0070/09 $26.00 © 2009 IEEE
lognormal distributions as another lognormal distribution by matching the mean and variance, which is referred to as Wilkinson’s approach. $P_{\text{chip}} = \sum_{i \in T} N_i P_i \approx e^{G}$, $\mu_G = \sum_{i \in T} N_i \mu_{Pi}$, and $\sigma_G^2 = \sum_{i \in T} N_i \sigma_{Pi}^2 + \sum_{i,j \in T} N_i N_j \text{cov}_{PiPj}$, where $G$ is a Gaussian random variable with mean $\mu_G$ and variance $\sigma_G^2$, and $\mu_{Pi}$ and $\sigma_{Pi}^2$ are the mean and variance of $P_i$, respectively. Wilkinson’s approach provides a closed-form approximation of chip-leakage-power distribution. For Wilkinson’s approach, in order to calculate full-chip leakage power, one needs to calculate the sum of covariances between each pair of circuit elements. Since there are $N_t$ types of circuit elements in the circuit and, for each pair of circuit elements, $n$ (number of variation sources) operations are required to calculate covariances, Wilkinson’s approach’s computational complexity is $O(nN_t^2)$, where $N_T = |T|$ is the number of circuit-element types (or the size of the standard cell library). However, Wilkinson’s approach is not directly additive, i.e., if we change one type of circuit element, the full-chip mean and variance need to be recalculated. Therefore, it is not very efficient for full-chip-leakage-power optimization.

III. ADDITIVE LEAKAGE-POWER MODEL

In this section, we introduce the additive leakage-power-variation model. The basic idea is to express leakage-power variation as a polynomial instead of an exponential function of variation sources, which is used in most of the statistical leakage analysis. Considering that leakage power is exponentially related to variation sources, in order to approximate leakage power accurately, we need high-order polynomials. In this paper, we approximate leakage power as a fourth-order polynomial of variation sources as follows:

$$P = P_0 \cdot \left( a_0 + \sum_{1 \leq i < j \leq n} b_{ij} X_i X_j + b_{ijk} X_i^2 X_j X_k \right) + \sum_{1 \leq i \leq n} \left( a_1 X_i + a_2 X_i^2 + a_3 X_i^3 + a_4 X_i^4 \right)$$

(1)

where $X = (X_1, X_2, \ldots, X_n)$ is the vector of variation sources and $a_{ik}, b_{ijk}$, and $a_{ij}$ are coefficients that can be obtained from curve fitting through SPICE simulation or measurement. In (1), we keep the first four order terms for each variation source and the first- and second-order pairwise crossing terms. There are $n^2 + 3n + 1$ terms in total in the polynomial, where $n$ is the number of variation sources. To achieve higher accuracy, users may use more terms at the cost of increased computational complexity.

With polynomial expression of leakage power, it is very easy to compute chip leakage power. Chip leakage power is calculated as the sum of the leakage powers of all circuit elements, $P_{\text{chip}} = \sum_{i \in T} N_i P_i$, where $T$ is the set of element types of the circuit and $N_i$ is the number of circuit elements of type $i$. Since the leakage powers of all circuit elements are expressed in a polynomial form as in (1), it is easy to see that $P_{\text{chip}}$ is also in the same form. However, we still need to obtain the distribution of chip leakage power. We approximate chip leakage as a lognormal random variable by matching the mean and variance [7]. That is $P \approx P_0 e^{R_N}$, where $R_N$ is a Gaussian random variable. Without loss of generality, assume that all variation sources are with zero mean and unit variance. From (1), it is easy to find that the mean of leakage power can be calculated as

$$\mu_P = E[P] = P_0 \left( a_0 + \sum_{1 \leq i < j \leq n} b_{ij} + \sum_{1 \leq i \leq n} (a_{1i} + a_{3i} m_{i3} + a_{4i} m_{i4}) \right)$$

(2)

$$m_{2P} = E[p^2] = P_0 \left( a_0^2 + \sum_{1 \leq i < j \leq n} (2a_{0i} a_{ij} + 2a_{0i} m_{i3} + 2a_{0i} m_{i4}) + a_{1i}^2 + a_{2i}^2 m_{i4} + a_{3i}^2 m_{i6} + a_{2i} a_{4i} m_{i8} + 2a_{1i} a_{2i} m_{i3} + 2a_{1i} m_{i3} m_{i4} + 2a_{1i} m_{i5} + 2a_{2i} m_{i5} m_{i4} + 2a_{2i} m_{i6} + a_{3i} m_{i4} m_{i7} \right)$$

(3)

$$\sigma_P^2 = m_{2P} - \mu_P^2$$

(4)

where $m_{i4}$ is the 4th-order moment for the $i$th variation source $X_i$. After obtaining the mean and variance of $P$, we may obtain the mean and variance of $R_N$ in the same way as [7], $\mu_N = 2 \log \mu_P - 0.5 \log (\mu_P^2 + \sigma_P^2) - \log P_0$ and $\sigma_N = \log (\mu_P^2 + \sigma_P^2) - 2 \log (\mu_P)$. With the aforementioned approximation, we may obtain the statistical characteristics of chip leakage power. For example, the Y% percentile point $W$ can be calculated as $W = \sqrt{2 \sigma_N} \text{erf}^{-1}(2Y - 1) + \mu_N$.

From the previous discussion, we see that, in order to compute the chip-leakage-power variation, we need to calculate the sum of $N_t$ polynomials with $K$ terms. Moreover, we also need to calculate the square of a $K$-term polynomial. Therefore, the total computational complexity is $O(K N_t + K^2)$. As discussed before, in this paper, we assume that $K = n^2 + 3n + 1$, which is the number of terms in (1). In practice, users may choose a different number of terms for the polynomial model. There is tradeoff between efficiency and accuracy. For example, we may approximate cell leakage power in a simpler or more complicated polynomial model as follows:

$$P = P_0 \cdot \left( a_0 + \sum_{1 \leq i < j \leq n} (b_{ij} X_i X_j) + \sum_{1 \leq i \leq n} \left( a_{1i} X_i + a_{2i} X_i^2 + a_{3i} X_i^3 \right) \right)$$

(5)

$$P = P_0 \cdot \left( a_0 + \sum_{1 \leq i < j \leq n} \left( b_{ij1} X_i X_j + b_{ij2} X_i^2 X_j^2 + b_{ij3} X_i^3 X_j^3 + b_{ij4} X_i^2 X_j X_k^2 \right) + \sum_{1 \leq i \leq n} \left( a_{1i} X_i + a_{2i} X_i^2 + a_{3i} X_i^3 + a_{4i} X_i^4 \right) \right)$$

(6)

1Here, we assume that all covariance terms are calculated and stored in a lookup table. Otherwise, the computational complexity of Wilkinson’s approach is even higher.
Equation (5) illustrates a simpler polynomial model that has only \(n(n-1)/2 + 3n + 1\) terms, while (6) illustrates a more complicated polynomial model that has \(5n(n-1)/2 + 5n + 1\) terms. Table I compares the error percentage of mean, variance, and 99% percentile point for the model in (1), (5), and (6). In the table, the average error percentage is calculated as the average of the absolute value. From the table, we find that accuracy increases as the number of polynomial terms increases, as expected. However, when the number of polynomial terms increases to a certain value as in the model of (6), the improvement of accuracy is limited. Therefore, in this paper, we will apply the model in (1).

On the other hand, the computational complexity for Wilkinson’s approach is \(N_n^2\), as discussed in Section II. Since the number of polynomial terms \(K\) is usually much smaller than the number of element types \(N_e\), our method will be more efficient than Wilkinson’s approach.

Moreover, during circuit optimization, fast incremental evaluation of power due to small circuit change is necessary. For Wilkinson’s approach, if one of the circuit elements is changed in the circuit, the sum of covariances between the changed circuit element and the remaining circuit elements needs to be recalculated; therefore, the computational complexity of recalculating chip leakage is \(O(N_e)\). However, for our method, we only need to calculate the sum of two \(O(K)\) polynomials and recalculate the mean and variance. In order to calculate the variance of a \(K\)-term polynomial, we need to calculate the mean of the square of it, which has \(K^2\) terms. Therefore, the computational complexity of recalculating chip leakage is \(O(K^2)\). That is, to recalculate chip leakage power, the computational complexity of our method does not depend on the number of circuit-element types (size of the standard cell library).

To show the efficiency of our approach, let us observe a simple example. Assume that we are going to calculate the leakage power of a circuit with 100 types of different circuit elements with two variation sources. For Wilkinson’s approach, it needs to calculate the sum of the covariances of each pair of types. There will be 100 \(\times (100 - 1)/2 = 4950\) pairs in total. However, for our approach, it only needs to calculate the sum of 100 polynomials with 11 terms. Moreover, if we want to incrementally calculate chip leakage power when one circuit element is changed, for Wilkinson’s approach, the sum of covariances between this circuit element and all other circuit elements needs to be calculated, and there will be 99 covariances. However, for our approach, we only need to calculate the sum of two 11-term polynomials.

Table I

<table>
<thead>
<tr>
<th>Method</th>
<th>Polynomial Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilkinson</td>
<td>(n(n-1)/2 + 3n + 1)</td>
</tr>
<tr>
<td>Equation (5)</td>
<td>(5n(n-1)/2 + 5n + 1)</td>
</tr>
</tbody>
</table>

In the previous section, we have introduced an additive polynomial leakage-power-variation model. Such a model works well when there is only interdie variation. In this section, we further discuss how to handle within-die variation in our model.

In practice, process variation is decomposed into interdie, within-die spatial, and within-die random variations \(X = X_g + X_s + X_r\), where \(X_g\) is the interdie variation, \(X_s\) is the within-die spatial variation, and \(X_r\) is the within-die random variation. Considering the within-die variation, the polynomial term \(X^k\) in (1) becomes \((X_g + X_s + X_r)^k\). Because \(X_g\) and \(X_s\) are different for different cells, such terms are not directly additive. In this case, we need to simplify the model to make it additive. First, we consider the within-die random variation. The polynomial term \(X^k\) can be calculated as

\[
X^k = (X_g + X_s + X_r)^k = \sum_{i=1}^{k} \binom{k}{i} (X_g + X_s)^{k-i}X_r^i. \tag{7}
\]

In practice, there are many cells of the same type with the same \(X_s\) and \(X_g\). Therefore, we need to compute the sum of the leakage powers of a large number of cells of the same type, and the \(X_s\)'s for different cells are independent. Therefore, due to the central limit theorem, similar to the method in [1], we may approximate the aforesaid equation as

\[
X^k \approx \sum_{i=1}^{k} \binom{k}{i} (X_g + X_s)^{k-i}E[X_r^i]. \tag{8}
\]

where \(E[X_r^i]\) can be obtained by the distribution of \(X_r\). In this case, within-die variation is modeled as the change of polynomial coefficients.

Second, let us consider spatial variation. Because \(X_s\)'s are correlated for different cells, we cannot apply the central limit theorem on it. In this case, we apply the grid-based spatial-variation model. That is, a chip is divided into several grids, and all cells within the same grid have the same spatial variation. The spatial variation of different grids is correlated. Moreover, by applying principal component analysis (PCA) on spatial variation, spatial variation can be expressed as a linear combination of principal components

\[
X_s = \sum p_i X_{pi}. \tag{9}
\]

Then, the terms \((X_g + X_s)^k = (X_g + \sum p_i X_{pi})^k\). We may regard \(X_{pi}\)'s and \(X_s\)'s as different variation sources and then obtain the polynomial expression by expanding \((X_g + \sum p_i X_{pi})^k\). One problem is that expanding \((X_g + \sum p_i X_{pi})^k\) results to a lot of terms and makes the polynomial very complex. However, considering the fact that spatial variation is usually not significant [8], [9], we may ignore the crossing terms of spatial variation without loss of much accuracy. Therefore, in the expansion, we may only keep the first three order terms of spatial variation and the first-order crossing terms between spatial variation and the corresponding interdie variation

\[
(X_g + \sum p_i X_{pi})^k \approx X_g + X_g^2 + X_g^3 + X_g^4 + \sum_{1 \leq i < n \leq k} (X_{pi} + X_{pi}^2 + X_{pi}^3 + kX_{pi}X_{pi}) \tag{9}
\]

where \(g_i\) is the number of grids (number of spatial-variation sources). With such approximation, in the polynomial expression of leakage power, we have \(n^2 + 3n + 1\) terms for interdie variation, as discussed in the previous section, and for each variation source, we have \(4g_i\) terms for within-die spatial variation and \(2g_i\) terms for the crossing terms between spatial and interdie variations. Therefore, the total

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on August 14,2010 at 02:12:30 UTC from IEEE Xplore. Restrictions apply.
number of variation terms is $K = n^2 + 3n + 1 + 4ng_s$. In the same way as that in the previous section, the computational complexity of full-chip-leakage-power calculation considering spatial variation is $O\left( (n^2 + ngs)N_t + (n^2 + ngs)^2 \right)$, and the computational complexity of incremental leakage-power calculation is $O\left( (n^2 + ngs)^2 \right)$. On the other hand, in this case, the computational complexity of Wilkinson’s approach is $O\left( ngsN_t^2 \right)$, and that of incremental leakage-power calculation is $O\left( ngsN_t \right)$. We can see that, when $N_t \gg g_s \gg n$, our method is $N_t$ times faster than Wilkinson’s approach. That is, the efficiency improvement of our method is even higher when we consider spatial variation.

Handling within-die variation needs to calculate the expansion of polynomials, which is time consuming. However, such an operation needs to be run only once for all types of cells. Therefore, it will not increase the computational complexity.

V. EXPERIMENTAL RESULTS

In the first set of experiments, we calculate leakage power for seven Microelectronics Center of North Carolina (MCNC) benchmarks [10]. For each benchmark, we implement it using a field-programmable gate-array (FPGA) circuit with lookup-table size $s = 4$. For variation analysis, we assume two variation sources, namely, channel gate length and dopant density. For each variation source, we consider both inter- and within-die variations. They follow a Gaussian distribution, and the $\sigma$ values of inter- and within-die variations are 8% and 6% of the nominal value, respectively. In our experiment, we assume an International Technology Roadmap for Semiconductors 32-nm high-performance technology and use a Model for Assessment of cmoS Technologies And Roadmaps-4 tool [11] to calculate leakage power. We also define two comparison cases: 1) Wilkinson’s approach and

![Fig. 1](image_url)

**Fig. 1.** (a) Run-time comparison for c1908 with varying numbers of cell types. (b) Estimated number of operations with varying numbers of cell types.
TABLE VI
MEAN, STANDARD-DEV., AND 99%-PERCENTILE-POINT COMPARISON FOR FULL-CHIP CALCULATION OF LEAKAGE POWER FOR ISCAS85
STANDARD CELL BENCHMARKS ASSUMING THE EXISTENCE OF SPATIAL CORRELATION

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>N_e</th>
<th>Method</th>
<th>Wilkinson</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>μ (mW)</td>
<td>σ (mW)</td>
<td>σ (mW)</td>
</tr>
<tr>
<td>c1355</td>
<td>39</td>
<td>6.89 (+0.0%)</td>
<td>1.73 (-3.9%)</td>
<td>10.12 (-4.0%)</td>
</tr>
<tr>
<td>c1908</td>
<td>32</td>
<td>10.16 (-0.4%)</td>
<td>2.15 (-1.8%)</td>
<td>15.75 (-0.8%)</td>
</tr>
<tr>
<td>c3540</td>
<td>46</td>
<td>21.32 (-0.9%)</td>
<td>4.08 (-2.8%)</td>
<td>28.91 (-1.6%)</td>
</tr>
<tr>
<td>c6288</td>
<td>43</td>
<td>32.11 (-0.3%)</td>
<td>6.86 (-2.3%)</td>
<td>45.35 (-1.7%)</td>
</tr>
<tr>
<td>c7552</td>
<td>45</td>
<td>46.84 (-0.2%)</td>
<td>10.69 (-3.9%)</td>
<td>69.42 (-1.8%)</td>
</tr>
<tr>
<td>ave</td>
<td>-</td>
<td>0.3%</td>
<td>3.1%</td>
<td>1.8%</td>
</tr>
</tbody>
</table>

Other than the aforementioned experiment, we have also done some experiments assuming the existence of spatial correlation. In the experiments, we assume that the 3σ values of interdie, within-die spatial, and within-die random variations are 8%, 4.2%, and 4.2% of the nominal value, respectively. For spatial variation, we apply the grid-based spatial-variation model in [12], and we divide each chip into four grids. Table VI shows the results of the full-chip leakage power with the existence of spatial correlation. From the table, we find that our approach is 22 times faster than Wilkinson’s approach with less than 2% accuracy loss. This validates the prior finding that the efficiency improvement of our method is higher when we consider spatial variation, as discussed in Section IV.

VI. CONCLUSION

The polynomial approximation of leakage power provides a decently simple, quick, and fairly accurate solution to calculate chip-leakage-power variation. Since only simple polynomial addition has been used, the method is very scalable, as well as incremental. It has been shown that, in order to recalculate chip-leakage-power variation after changing some circuit elements, the computational complexity of our approach does not depend on the number of circuit-element types (standard-cell-library size), while the complexity of Wilkinson’s approach, which is widely used in many existing statistical leakage analysis works, is linear to the circuit-element types. Experimental results have shown that our method is five times more efficient than Wilkinson’s approach with no accuracy loss in mean estimation and about 1% accuracy loss in standard-deviation and 99%-percentile-point estimations.

ACKNOWLEDGMENT

The authors would thank R. Bell for the helpful discussion and initial work.

REFERENCES


Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on August 14,2010 at 02:12:30 UTC from IEEE Xplore. Restrictions apply.