Robust Estimation of Parametric Yield under Limited Descriptions of Uncertainty

Wei-Shen Wang and Michael Orshansky
Department of Electrical and Computer Engineering
The University of Texas at Austin

ABSTRACT

Reliable prediction of parametric yield for a specific design is difficult; a significant reason is the reliance of the yield estimation methods on the hard-to-measure distributional properties of the process data. Existing methods are inadequate when dealing with real-life distributions of process and environmental parameters, and limited availability of parameter data during early design. This paper proposes a robust technique for full-chip parametric yield estimation; the proposed work is based on the rigorous notions of non-parametric robust statistics which permits estimation based on the knowledge of the range and the limited number of moments (e.g. mean and variance) of the parameter distributions. Fully or partially specified process and environmental parameters can be described by robust representations, and used to estimate probabilistic bounds for leakage dissipation. The proposed approach is applied to estimating the chip-level parametric yield. The experimental results show that the robust estimation algorithm improves the total leakage estimate by 5-13% at the 99th percentile across distinct frequency bins, compared to using only the intervals of partially-specified parameters.

1. INTRODUCTION

The growth of standby, or leakage, power as device geometries scale down has been acknowledged as a major potential threat to scaling. At the 65nm node, leakage power may account for almost half of total power of the circuit [1]. The exponential dependence of leakage on some process parameters also causes a large spread in leakage current in the presence of process variations: a 30% variation in the effective channel length could potentially lead to 20X variation in leakage current [2]. Given the magnitude of leakage variation, quantifying its impact on parametric yield is crucial for early chip design [3].

The technique for joint timing- and power-limited parametric yield estimation of [3] relied on the fact that yield is limited both by leakage power consumption and chip frequency. Leakage power is inversely correlated with chip frequency; slow die have low leakage, while fast die have high leakage. The exponential dependence of leakage on process spread means that the total power will cross the cooling (power) limit well below the maximum possible chip frequency since chips operating at higher frequencies have exponentially higher leakage power consumption. A gate-level parametric yield estimation algorithm was proposed in [4]. In addition, several papers separately studied the statistical leakage estimation problem. In [5], a gate-level full-chip leakage analysis algorithm taking into account spatial correlations of intra-chip process variations was proposed. In [2], a probabilistic approach was proposed to estimate subthreshold leakage distribution accounting for intra-chip and inter-chip variations of process parameters, temperature, and supply voltage. All the above techniques, however, rely on idealized assumptions about variability of process and environmental parameters.

Practical application of statistical leakage and parametric yield analysis techniques is severely limited by fundamental features of a real-life industrial IC design process. First, process characterization data is often incomplete due to the limited number of measurements and characterization lots. As a result, there may be a large uncertainty in the statistic metrics (the mean and the variance of process parameters). Secondly, the physical mechanisms responsible for generating variability are often not understood well, especially for new technologies that exhibit physical behaviors not encountered. As a consequence, the typical assumption of a normal distribution may be incorrect. There are parameters that behave decidedly non-normally, such as via resistance, yet its precise distribution is difficult to model. This requires the use of non-Gaussian models for accurate estimation, which cannot be handled by existing analytical methods.

This paper addresses the limitations of the existing methods by introducing a new mathematical formulation that enables robust prediction of timing- and power-limited parametric yield. Here, robustness is defined as the insensitivity of the produced estimates to the idealized modeling assumptions of earlier techniques. The fundamental theory behind the work is probabilistic interval analysis that extends the representation of a random variable to a family of distributions, i.e., bounds for cumulative distribution functions, and thus can work with a wider class of uncertainty models. This paradigm of handling partially-specified uncertainty has been applied to timing analysis [6], and effectively reduces the over-conservatism of the interval-based prediction. In contrast to pure interval analysis and affine methods, probabilistic interval analysis has the capabilities of preserving the notion of probability. In affine methods, the notion of probability has to be recovered heuristically [7].

The proposed algorithm is based on non-parametric robust statistics, which permits using statistic metrics (e.g. the mean and variance) to describe partially-specified environmental fluctuations in chips. Probabilistic arithmetic based on linear programming [8] is used to compute probabilistic bounds for functions of random variables. This strategy, along with realistic modeling of process variability, is able to assess the impact of process and environmental variations on leakage dissipation and estimate the guaranteed parametric yield. The experimental results indicate that at the 99th percentile the proposed strategy improves the estimated total leakage current by 5-13% across the $\pm 3\sigma$ range of the inter-chip effective channel length variation. Thus, the proposed work utilizes the partial probabilistic descriptions to effectively improve the leakage estimate obtained

ICCAD'06, November 5–9, 2006, San Jose, CA

from using only the interval information of partially-specified environmental parameters.

The paper is organized as follows. Section 2 describes the needs for robust estimation methodologies. In Section 3, the mathematical formulation of probabilistic interval methods is presented. The full-chip leakage modeling and yield estimation are described in Section 4. In Section 5, the experimental results are presented.

2. NEED FOR ROBUST ESTIMATION

Fundamentally, robust methods enable the use of partial probabilistic information with an interval-based uncertainty representation. Traditional statistical leakage and yield estimation techniques have assumed that complete distribution information about the random parameters is available, and random variables belong to the classical single-mode parametric distributions (e.g. normal) generated by a unique stochastic process. In practice, these assumptions often do not hold.

(i) Incomplete Probabilistic Information: For some sources of variability, the probabilistic information is either incomplete. For example, variability in supply voltage and on-chip temperature affects leakage super-linearly [2]. An accurate estimate could be produced by characterizing the distribution across the temporal and spatial domains. However, due to the computational difficulty of performing temporal input-dependent analysis, the uncertainty about supply voltage is typically represented by the range information [9].

Incomplete probabilistic information can be used to improve the quality of estimates, in some cases. Although a complete characterization of power supply voltage or on-chip temperature distribution is virtually impossible, the mean and the variance can be easily estimated by a Monte Carlo sampling procedure. Besides, the variability of temperature can be estimated using logic-level temperature estimators [10]. Consequently, it is possible to estimate moments of environmental parameters, which permits assessing the impact of these factors on leakage.

(ii) Limitations of Parametric Modeling Strategies: Working with idealized models often requires adopting unreasonable assumptions about the dependence on process parameters. In the literature, an empirical model of the following form was found to accurately model leakage dependence on the key process parameters that are subject to substantial variability [11]:

$$I_{sub} = I_o W \exp[a_1 + a_2 L + a_3 L^2 + a_4 T_{ox}^{-1} + a_5 T_{ox}] \quad (1)$$

where L is the effective channel length, and T_{ox} is the oxide thickness. Then, in order to fit the distribution of I_{sub} to a log-normal distribution, (1) is simplified into:

$$I_{sub} = I_{nom} W \exp[b_0 + b_1 \Delta L + b_2 \Delta T_{ox}] \quad (2)$$

Given that the inter-chip variation of L is significant, and because of the exponential effect that an approximation will have on the result, this transformation may not be acceptable. The cost of not performing such a transformation is that I_{sub} is not characterized by a lognormal distribution, which poses difficulty for existing leakage analysis methods. We will solve this problem by adopting self-verifying robust methods of estimation. Besides, the current process technology is well-controlled; the outlier of parameters is unlikely to exist in fabricated chips. Therefore, truncated distributions are more appropriate to represent process variability because it avoids the erroneous estimation resulting from the infinite tails of Gaussian variables.

(iii) Non-Gaussian, Mixture, and Multi-Modal Distributions: In some cases, the distributions of parameters exhibit non-Gaussian, mixture [12], or multi-modal behavior, i.e., the probability density function has multiple peaks. Algorithms based on parametric techniques are notoriously poor at handling these distributions; the distributions can only be approximated as parametric distributions (e.g. normal) for convenient manipulations. However, the robust estimation framework can naturally handle a variety of distributions including non-Gaussian, mixture, and multi-modal variables, as demonstrated in Section 3.

An extreme case for the mixture distribution is when a fabless company works with two or more foundries in manufacturing a design. The design must be robust under distinct fab-specific parameter distributions. The formal statistical model to analyze this case is finite mixture distribution [12]. Suppose chips are fabricated by n (e.g. n=2) fabs. The probability density function (*pdf*) of effective channel length, $f(L)$, can be computed by:

$$f(L) = \sum_{i=1}^{n} f(L \mid F_i) \cdot p(F_i) \quad (3)$$

where $p(F_i)$ is the probability that a chip is fabricated in fab i, and $f(L \mid F_i)$ is the conditional *pdf* for fab i. The mean (μ) and variance (σ^2) of the mixture distribution can be computed by:

$$\mu = \sum_{i=1}^{n} p(F_i)\mu_i \qquad \sigma^2 = \sum_{i=1}^{n} p(F_i)\left(\mu_i^2 + \sigma_i^2\right) - \mu^2 \quad (4)$$

where μ_i and σ_i^2 are the mean and variance of L from fab i.

Figure 1 shows the mixture of the effective channel length distributions due to statistically distinct populations, assuming each foundry contributes half of the manufactured chips. Approximating the mixture distribution as a normal distribution causes inaccurate modeling of channel length variations, thus overestimates the subthreshold leakage of a transistor by 13% and 48%, at the 95th and 99th percentile, respectively. This example points out the limitation of the analytical approaches based on parametric techniques and idealized assumptions about variability.

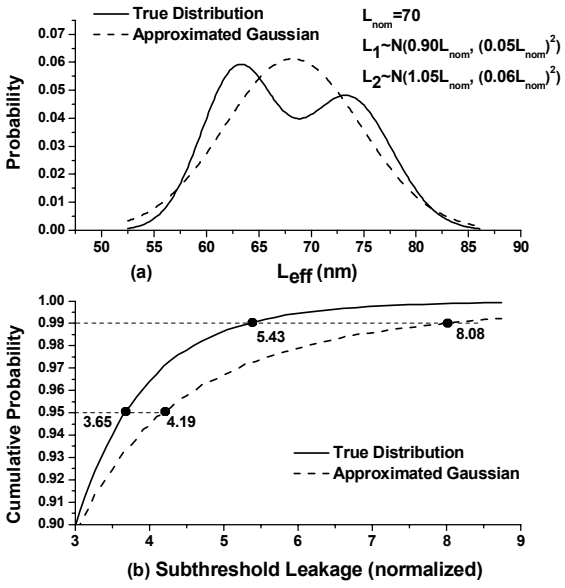

Figure 1. Approximating uncertainty as a Gaussian variable may lead to a large error in leakage estimation (a) channel length distribution (b) subthreshold leakage distribution.

3. FUNDAMENTAL PHILOSOPHY OF ROBUST COMPUTATION OF RANDOM VARIABLES

In this section a formal description of robust estimation procedure is given. Its purpose is to enable *reliable* and *assumption-free* generation of distributions of functions of random variables. The adopted framework can be seen as a probabilistic interval method. It supplements the estimates of interval and affine methods with the partial probabilistic information enabling a new type of analysis. The framework requires the development of two distinct sets of mathematical tools for robust representation of random variables, and robust operations with random variables.

3.1 Robust Representation of Random Variables

In a robust estimation framework, an appropriate representation of uncertainty is needed. Conceptually, the most general representation of a fully specified random variable is its cumulative distribution function (*cdf*) [13]. For a partially-specified random variable, the most general representation is a set of cumulative distribution functions, or a *p-box* [14].

Definition: $\overline{F}$ and $\underline{F}$ are non-decreasing functions from $\Re$ into [0, 1], and $\underline{F}(x) \le \overline{F}(x), x \in \Re$. A *p-box*, denoted by $[\underline{F}, \overline{F}]$, is defined as a set of imprecisely known cumulative distribution functions, $F(x) = P(X \le x)$, where $\underline{F}(x) \le F(x) \le \overline{F}(x)$.

A p-box represents upper and lower bounds for the cumulative distribution function of a random variable. It is a basic notion for the robust computation, and can be used to robustly describe a random variable. Because the p-box representation is parametric-free, it can be used to describe a variety of distributions including the non-Gaussian, multi-modal, and mixture distributions described in Section 2. For instance, given the cumulative distribution function of a random variable, $F(x)$, we can sample it at a series of non-decreasing values, x_i, where $F(x_0) = 0$ and $F(x_n) = 1$, $0 \le i \le n$. Then a p-box can be constructed as:

$$\begin{aligned} &\overline{F}(x) = F(x_{i+1}) \\ &\underline{F}(x) = F(x_i) && x_i \le x < x_{i+1},\ i = 0..n-1. \\ &\overline{F}(x) = \underline{F}(x) = F(x) && x \ge x_n \text{ or } x < x_0 \end{aligned}$$

Thus, the p-box representation enables our framework to incorporate true distributions of process variability, instead of resorting to the normal assumption of parameters.

Another useful description of uncertainty is in terms of intervals with *partial probabilistic information.* In addition to the bounds, often limited information, such as the mean or the variance, is available. In this situation, it seems wasteful not to use this information. From Section 2, these statistic metrics of environmental parameters can be estimated during the early design phase; therefore, these parameters are modeled as probabilistic intervals in our framework. Besides, probabilistic intervals can be used to describe partially-specified process parameters if statistic metrics and range are available.

The *probabilistic interval* description needs to be converted into a p-box representation for further manipulations. This is done using a sophisticated generalization of the one-sided Chebyshev inequality [13] and Cantelli inequality [15] which enables computing bounds of the cumulative probability [14]. The upper bound for the cumulative probability of a random variable X is:

$$\begin{aligned} &P(X \le x) = 0 && x < \underline{X} \\ &P(X \le x) \le 1 \Big/ \left(1 + (\mu - x)^2 / \sigma^2\right) && \underline{X} \le x < \mu + \sigma^2/(\mu - \overline{X}) \\ &P(X \le x) \le 1 - \left(m^2 - my + s^2\right)/(1 - y) && \mu + \sigma^2/(\mu - \overline{X}) \le x < \mu + \sigma^2/(\mu - \underline{X}) \\ &P(X \le x) = 1 && \mu + \sigma^2/(\mu - \underline{X}) \le x \end{aligned}$$

where $\underline{X}$ and $\overline{X}$ denote lower and upper bounds, μ and σ^2 denote the mean and variance, $y = (x - \underline{X})/(\overline{X} - \underline{X})$, $m = (\mu - \underline{X})/(\overline{X} - \underline{X})$, and $s^2 = \sigma^2/(\overline{X} - \underline{X})^2$. Similarly, the lower bound is given by:

$$\begin{aligned} &P(X \le x) = 0 && x < \mu + \sigma^2/(\mu - \overline{X}) \\ &P(X \le x) \ge 1 - (m(1+y) - s^2 - m^2)/y && \mu + \sigma^2/(\mu - \overline{X}) \le x < \mu + \sigma^2/(\mu - \underline{X}) \\ &P(X \le x) \ge 1/(1 + \sigma^2/(x - \mu)^2) && \mu + \sigma^2/(\mu - \underline{X}) \le x < \overline{X} \\ &P(X \le x) = 1 && \overline{X} \le x \end{aligned}$$

Figure 2 shows a p-box for the uncertainty of power supply, which in fact represents all distributions with the same mean, variance, and range. The p-box permits estimating the uncertainty at any confidence level. For example, in Figure 2 when the cumulative probability is 0.50, the right-side p-box falls at -0.02V, which means at least 50% of the samples have ΔV_{dd} less than or equal to -0.02V, i.e., $P(\Delta V_{dd} \le -0.02\text{V}) \ge 0.50$. Thus, we can easily estimate the percentage of the samples meeting a specific requirement using p-boxes.

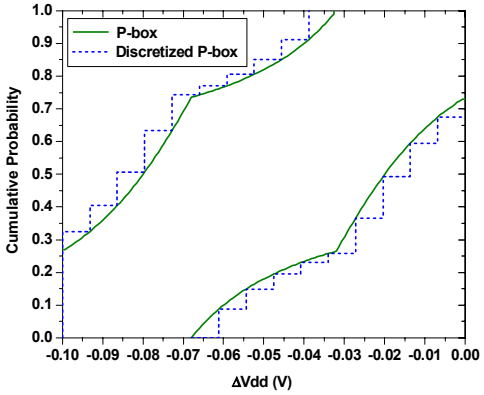

Figure 2. The knowledge of range, mean and variance permits constructing a p-box for a variable. The mean and the variance values are -0.05V and $(0.03\text{V})^2$, respectively.

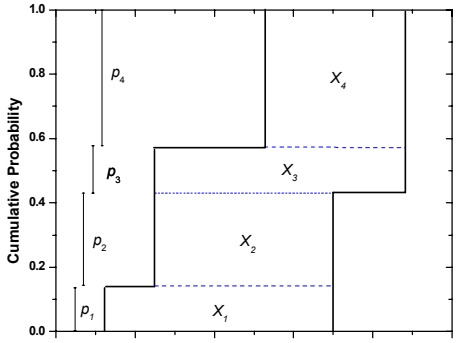

Figure 3. Transformation of a discretized p-box into a histogram representation.

The p-box representations described above are useful for describing process and environmental parameters. The robust estimation framework seeks to perform numerical operations on p-box descriptions of variables, and provide guaranteed computational results (e.g. total leakage dissipation) that are also described by p-boxes. In order to implement computation with p-boxes, an intermediate and numerically tractable representation is needed during robust computations. This is based on the notion of *self-validating histograms* [8].

Definition: A *self-validating histogram* of a random variable X is:

$$X = \bigcup_i X_i \ , \quad X_i = \left[\underline{X_i}, \overline{X_i}\right]$$
$$P\left(X \in X_i\right) = p_i \text{ for all } i, \text{ and } \sum_i p_i = 1 .$$

where X_i is an interval associated with the probability p_i.

This histogram representation describes a random variable as a set of intervals associated with probabilities. As a *two-valued* histogram, in which lower and upper endpoints of the intervals are recorded, the histogram is self-validating because it is able to keep track of the accuracy (error) of the computed quantities. Before numerical computations of p-boxes, we need to transform p-boxes into histograms [16], which will be clear in Section 3.2. This transformation requires two phases: first, the p-box needs to be conservatively discretized as in Figure 2, which means that the discretized p-box forms an envelope of the original p-box. The discretization can be done with arbitrary granularity depending on the required accuracy. Then the discretized p-box is transformed into the two-valued histogram, as shown in Figure 3. Note that the intervals, X_i, may overlap, which provide great flexibility of describing random variables.

Having transformed random variables into the self-validating histograms, we can then perform arithmetic operations.

3.2 Robust Operations with Uncertain Variables

Various arithmetic operations can be performed on variables described by self-validating histograms, establishing general probabilistic arithmetic [8][17], which permits computing the p-box for functions of random variables. In this section, we demonstrate the arithmetic operations on single, and multiple random variables described by self-validating histograms.

First we describe how to evaluate arbitrary functions of a random variable X, given the histogram representation. This computation can be done by creating a table including all intervals, as in Figure 4. For each interval of X, we compute upper and lower bounds of the function $Z = f(X)$ when $X \in [\underline{X_i}, \overline{X_i}]$. The bounds are

$$\underline{Z_i} = \min f\left(X \in [\underline{X_i}, \overline{X_i}]\right), \ \overline{Z_i} = \max f\left(X \in [\underline{X_i}, \overline{X_i}]\right),$$

where $\underline{Z_i}$ and $\overline{Z_i}$ denotes the lower and upper bounds of the function, respectively, when $X \in [\underline{X_i}, \overline{X_i}]$.

It is apparent that $P(X \in [\underline{X_i}, \overline{X_i}]) = P(Z \in [\underline{Z_i}, \overline{Z_i}])$. Thus, a histogram of $f(X)$ is constructed. The final step is to compute the p-box of $f(X)$, i.e., the bounds for the *cdf*. The cumulative probability of the function at a specific value z, $P(Z \le z)$, is actually bounded by:

$\boldsymbol{X}$	$\boldsymbol{Z = f(X)}$
⋮	⋮
$X \in [\underline{X_i}, \overline{X_i}]$	$Z \in [\underline{Z_i}, \overline{Z_i}],\ p_i = P(Z \in [\underline{Z_i}, \overline{Z_i}])$
⋮	⋮

Figure 4. Probability table for a function of a random variable.

$\boldsymbol{Z = f(X,Y)}$	…	$Y \in [\underline{Y_j}, \overline{Y_j}]$	…
⋮	⋱	⋮	⋰
$X \in [\underline{X_i}, \overline{X_i}]$	…	$Z \in [\underline{Z_{ij}}, \overline{Z_{ij}}],\ p_{ij} = P(Z \in [\underline{Z_{ij}}, \overline{Z_{ij}}])$	…
⋮	⋰	…	⋱

Figure 5. Probability table for a function of multiple random variables.

$$\begin{aligned} P(Z \le z) &\le \textstyle\sum_i p_i \quad \forall i : \underline{Z_i} \le z \\ P(Z \le z) &\ge \textstyle\sum_i p_i \quad \forall i : \overline{Z_i} \le z \end{aligned} \tag{5}$$

That is, when computing the upper bound of $P(Z \le z)$, the intervals with the lower bound $\underline{Z_i}$ no larger than z need to be considered. In the histogram representation, the probability mass of an interval is specified, i.e., $p_i = P(Z \in [\underline{Z_i}, \overline{Z_i}])$; however, how the probability mass is distributed within the interval is not constrained. When the probability mass p_i of an interval entirely falls at the lower bound of the interval $\underline{Z_i}$, i.e., $p_i = P(Z = \underline{Z_i})$, it results in the largest increase in the *cdf*, thus determines the upper bound of the *cdf*. Similarly, the lower bound of the *cdf* can be determined when the probability falls at the upper bound of the interval. As a result, we can use (5) to compute the p-box of $f(X)$. Besides, the bounds for statistic metrics can be computed from the histogram representation. For instance, the bounds for the expected value of Z are given by:

$$\underline{E}[Z] = \textstyle\sum_i p_i \underline{Z_i} \ , \ \overline{E}[Z] = \sum_i p_i \overline{Z_i} \tag{6}$$

where $\underline{E}[Z]$ and $\overline{E}[Z]$ denote the lower and upper bounds for $E[Z]$.

For operations on multiple variables, the computation of p-boxes is an optimization problem because the distribution of the result depends on the correlation of variables. From [17], the operation of multiple random variables described by histograms requires solving a linear optimization problem. We briefly describe how to compute the bound for a function of two variables, X and Y, with unknown dependency.

First a discrete two-dimensional table is constructed to include all combinations of intervals from X and Y. For each cell we compute the bounds of the function, $f(X,Y)$.

$$\underline{Z_{ij}} = \min f\left(X \in X_i, Y \in Y_j\right), \ \overline{Z_{ij}} = \max f\left(X \in X_i, Y \in Y_j\right).$$

Then we assign a variable, p, as the probability mass for the cell.

$$p_{ij} = P(Z \in [\underline{Z_{ij}}, \overline{Z_{ij}}])$$

The constructed table is shown in Figure 5. Similar to (5), the cumulative probability of the function, $P(Z \le z)$, is bounded by:

$$\begin{aligned} P(Z \le z) &\le \textstyle\sum_{i,j} p_{ij} \quad \forall i,j : \underline{Z_{ij}} \le z \\ P(Z \le z) &\ge \textstyle\sum_{i,j} p_{ij} \quad \forall i,j : \overline{Z_{ij}} \le z \end{aligned} \tag{7}$$

Note that the sum of probabilities of cells in the same row (column) should be equal to the marginal probability of X (Y). Thus we have constraints for the variable, p_{ij}. The p-box of the multivariate function can be then computed by solving the optimization problems below for distinct values of z.

(i) The upper bound of the cumulative probability:

$$\max \quad \sum_{i,j} p_{ij} \quad \forall i,j : \underline{Z_{ij}} \le z .$$

(ii) The lower bound of the cumulative probability:

$$\min \quad \sum_{i,j} p_{ij} \quad \forall i,j : \overline{Z_{ij}} \le z .$$

The constraints of the optimization problems are:

$$\sum_i p_{ij} = P(Y \in [\underline{Y_j}, \overline{Y_j}]) \quad \text{for all } j.$$
$$\sum_j p_{ij} = P(X \in [\underline{X_i}, \overline{X_i}]) \quad \text{for all } i. .$$
$$p_{ij} \ge 0 \quad \text{for all } i,j.$$

Since the objective function and all constraints are linear functions of the cell probability, p_{ij}, the optimization problem is a linear programming problem that can be solved efficiently. Besides, this probabilistic arithmetic can be extended to handle multiple variables by manipulating a multi-dimensional probability table. Consequently, we are able to compute the p-box of any arbitrary function of variables described by the histogram representations.

A special case for operations on random variables is that variables are mutually independent. In this situation, the result of arithmetic operations can be evaluated without solving the optimization problem. For example, consider two independent random variables X and Y, the joint table is a two-dimensional grid, in which the probability of the entries is generated by

$$P(X \in X_i, Y \in Y_j) = P(X \in X_i) \cdot P(Y \in Y_j) . \qquad (8)$$

Once the associated probability of each cell is computed, we are able to construct the histogram of any function of probabilistic interval variables. With the histogram, the probability bound can be evaluated using (7).

4. ROBUST PARAMETRIC YIELD ESTIMATION

Reliably estimating parametric yield is extremely important for chip designers. The two-sided squeeze on yield means that yield estimation essentially requires reliable frequency and leakage prediction. We apply the robust estimation framework developed above to the problem of reliably evaluating the chip-level parametric yield. Since the input of the chip-level problem is relatively small, the computational cost is not a concern in this work. All factors that affect the robustness of yield prediction, described in Section 2, are included in the analysis. The variability of three process parameters is considered: effective channel length (L), threshold voltage (V_{th}), and oxide thickness (T_{ox}).

The subthreshold leakage and gate current models adopted here [1][3], describes it as an exponential function of the effective channel length, subthreshold voltage, and power supply voltage (V_{dd}). The dependency to on-chip temperature (T) is super-linear [2]; however, the leakage can be well approximated as an exponential function according to SPICE simulations. Therefore, similar to [1][3], the subthreshold leakage current of a unit-width transistor, is modeled as:

$$I_{sub} = I_{sub,nom} \cdot e^{a\Delta L_l^2 + (2a\Delta L_g + b)\Delta L_l + c\Delta V_{th,l} + d\Delta V_{dd} + e\Delta T} \cdot e^{a\Delta L_g^2 + b\Delta L_g + c\Delta V_{th,g}} \quad (9)$$

where $I_{sub,nom}$ is the nominal value of the subthreshold leakage current, $(\Delta L_l, \Delta V_{th,l})$ and $(\Delta L_g, \Delta V_{th,g})$ denote intra-chip and inter-chip components of variation. This model can effectively describe the quadratic dependency of the exponent on L variability.

The purpose of this work is to estimate the parametric yield for each frequency bin. Following [3], it is assumed that the inter-chip channel length variation (ΔL_g) largely determines the frequency, and thus can be assumed fixed for a specific bin.

Now we are capable of computing the cumulative probability of I_{sub}. In this framework, environmental fluctuations (V_{dd} and T) are represented by probabilistic intervals, described in Section 3.1. For the purpose of demonstration, process variability (L and V_{th}) is modeled as a truncated Gaussian variable; however, our robust estimation is able to handle various distributions because fully or partially specified uncertainty can be described by the p-box representations.

Assuming variability due to distinct categories of parameters is independent, we can construct a multidimensional table for I_{sub} using (8), and then compute the interval and the probability for each cell. Since I_{sub} is a monotonic function of all the parameters of interest, its range can be efficiently computed by considering the combinations of endpoints for parameters within each cell. Finally, we obtain a histogram representation of I_{sub}. The p-box representation can be then computed using (7).

The total subthreshold leakage is computed by summing up the contribution of all transistors on the chip.

$$I_{sub,total} = \sum_i W_i I_{sub,nom} e^{a\Delta L_l^2 + (2a\Delta L_g + b)\Delta L_l + c\Delta V_{th,l} + d\Delta V_{dd} + e\Delta T} e^{a\Delta L_g^2 + b\Delta L_g + c\Delta V_{th,g}}$$

where W_i denotes the equivalent device width accounting for complex gates and stack effects of leakage. Because all transistors share the inter-chip component of variation, we can assess the impact of the intra-chip variation first, which can be evaluated by estimating the mean of the intra-chip variation as in [3]:

$$I_{sub,total} = I_{sub,nom} e^{a\Delta L_g^2 + b\Delta L_g + c\Delta V_{th,g}} \lambda_{sub} \sum_i W_i \qquad (10)$$

where $\lambda_{sub} = E[e^{a\Delta L_l^2 + (2a\Delta L_g + b)\Delta L_l + c\Delta V_{th,l} + d\Delta V_{dd} + e\Delta T}]$. The computation of the intra-chip factor λ_{sub} can be done by creating a multidimensional probability table and a histogram, which accounts for the uncertainty of $(\Delta L_l, \Delta V_{th,l}, \Delta V_{dd}, \Delta T)$. The lower (upper) bound for the mean of the histogram representation are computed using (6). Thus, we obtain the bounds of λ_{sub}.

For a chip-level analysis, the impact of environmental parameters can be evaluated using the statistic metrics (e.g. mean and variance) across the entire chip. If the statistic metrics of individual blocks in the chip design are available, however, our robust estimation framework can utilize these block-specific descriptions, and provide accurate leakage estimates. For example, a chip design based on the voltage island paradigm [18] may have distinct profiles of environmental parameters for blocks. Suppose the block-level statistic metrics of environmental parameters are available, our robust estimation framework is able to evaluate $\lambda_{sub} \sum_i W_i$ for each block, and sum up this term for all blocks on the chip. After evaluating the intra-chip factor and the equivalent width in (10), we can then construct a histogram for $\Delta V_{th,g}$, compute the histogram of $I_{sub,total}$, and obtain the p-box of the total subthreshold leakage current.

Similarly, we can compute bounds on the intra-chip gate leakage distribution over T_{ox} and V_{dd}. The total gate leakage current is expressed as:

$$I_{gate,total} = I_{gate,nom} e^{h\Delta T_{ox,g}} E\left[e^{h\Delta T_{ox,i}+k\Delta V_{dd}}\right]\sum_i W_i \qquad (11)$$

This model captures the dependency of gate leakage on T_{ox} and V_{dd}, and insensitivity to the on-chip temperature [19].

The first step of parametric yield evaluation is to find the distribution of the total leakage current which is the sum of the gate and subthreshold leakage sources. The previous methods [3] have assumed independence of subthreshold and gate leakages. In our model, however, subthreshold and gate leakage currents are correlated due to the dependence on variability of V_{dd}. As a result, the probability of the sum cannot be computed by convolution of individual *pdfs*; this sum of leakage currents must be computed by the probabilistic arithmetic described in Section 3, which can handle random variables with arbitrary correlations.

Finally, for every fixed value of the inter-chip channel length variation (ΔL_g), we use the abovementioned algorithm to compute the p-box of the total leakage current. Then we are able to compute the parametric yield for all frequency bins.

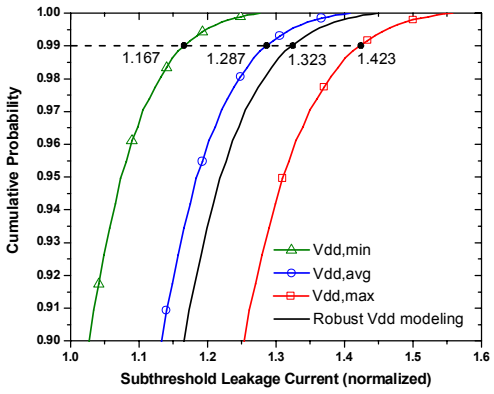

Figure 6. Total subthreshold leakage considering process variability (L and V_{th}) and V_{dd} uncertainty $(\Delta L_g = 0)$. The robust method improves the estimate based on the maximum V_{dd} by 7.0% at the 99th percentile.

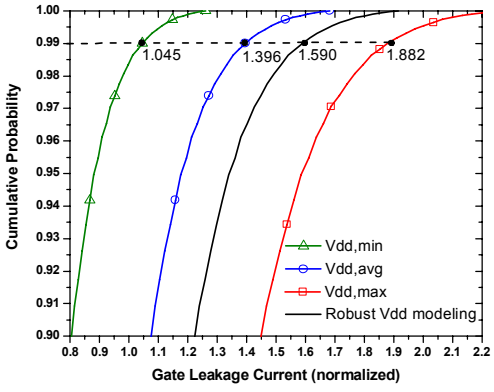

Figure 7. Total gate leakage considering process variability (T_{ox}) and V_{dd} uncertainty. The robust estimation strategy improves the estimate based on the maximum V_{dd} by 15.5% at the 99th percentile.

5. EXPERIMENTAL RESULTS

The robust nature of the proposed strategy allows us to compute bounds for the cumulative probability of the leakage current, instead of an approximated value. Since our primary objective is to estimate the guaranteed parametric yield, we focus on the lower bound of the cumulative probability, i.e., the upper bound of the leakage dissipation at any percentile. Thus, only the right-side p-box is shown in all figures. In the experiments, coefficients of parameters in subthreshold and gate leakage modeling are obtained from SPICE simulations for devices of PTM 70nm technology [20][21]. The 3σ values of L, V_{th}, and T_{ox} parameters are 20%, 10%, and 8% of the nominal values, respectively, with 50% of the variance contributed by the inter-chip component. The target chip is divided into 16 blocks with distinct ranges of environmental parameters; the maximum voltage drop is about 10-12% of the nominal value, and the standard deviation is about 3%. The range of on-chip temperature spans about 20°, with the standard deviation about 3°.

Figure 6 illustrates the importance of taking into account the uncertainty of supply voltage. We compute the lower bound for the *cdf* of the total subthreshold leakage current for zero inter-chip channel length variation $(\Delta L_g = 0)$. To evaluate the impact of V_{dd} uncertainty, we also compute the *cdf* assuming V_{dd} is a fixed value. The minimum, maximum, and average V_{dd} values of blocks are used. The robust estimation strategy predicts that the total subthreshold leakage is 1.323X of the nominal value at the 99th percentile, which means that at least 99% of the samples have leakage current no larger than 1.323X of the nominal case, i.e., $P(I_{sub,total} \le 1.323 I_{sub,total,nom}) \ge 0.99$. In contrast, the leakage estimates based on the maximum and minimum V_{dd} are 1.423X and 1.167X at the 99th percentile. Thus, the robust strategy improves the leakage estimate by 7.0%, compared to using the maximum V_{dd}. Similarly, the robust strategy improves the gate leakage dissipation by 15.5% at the 99th percentile, as illustrated in Figure 7. Consequently, the robust method can utilize the partial probabilistic descriptions to enhance pure interval analysis of environmental parameters. Additionally, Figure 6 illustrates that it is inappropriate to use the average value of power supply voltage because it predicts a lower estimate of leakage consumption, thus results in an over-optimistic prediction of parametric yield.

We also compare our robust methodology with prior work on leakage analysis, e.g. [3]. Because the algorithm in [3] does not take into account supply voltage and on-chip temperature, these parameters are assumed to be fixed values. Now the maximum values of supply voltage and temperature are incorporated in the algorithm described in [3] because the objective is to provide a guaranteed parametric yield. Figure 8 shows the total leakage dissipation computed by both approaches for a frequency bin $(\Delta L_g = 0)$. Compared to [3], our estimation approach provides a lower leakage estimate at any percentile due to robust modeling of environmental parameters; the robust strategy improves the total leakage estimates by 11.0% and 9.5% at the 50th and 99th percentiles, respectively. Besides, the robust estimation method predicts a higher parametric yield for a given limit of leakage current. Figure 9 shows the equi-yield contours for total leakage dissipation computed by both approaches. The difference in the 99th-percentile leakage consumption ranges from 5.3% to 13.4% within the $\pm 3\sigma$ range of the inter-chip channel length variation, as shown in Table I. Note that the difference between the contours of the same yield becomes pronounced for chips with the negative inter-chip L variation. Therefore, for chips with large leakage

currents, the robust approach can provide a more accurate estimation of parametric yield. It helps designers save extra efforts to apply additional leakage reduction techniques, and validates the necessity of adopting a robust estimation approach.

6. CONCLUSIONS

In this work a robust estimation approach is proposed to compute the chip-level parametric yield. Based on robust representations and operations of random variables, the proposed strategy is able to manipulate a variety of distributions that cannot be handled by analytical techniques, and takes into account the correlation of variables. Given statistic metrics of environmental parameters, the robust estimation methodology is able to provide guaranteed bounds for parametric yield, thus improve the estimate obtained from using only the intervals of environmental parameters.

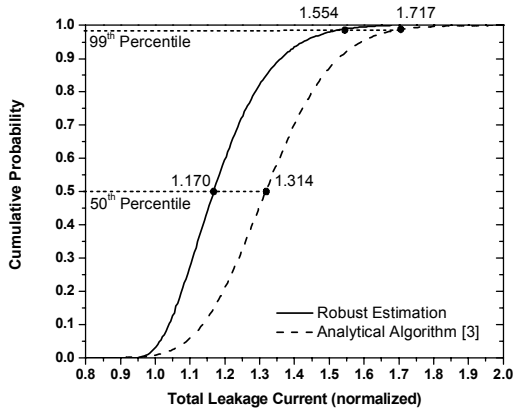

Figure 8. Total leakage current $(\Delta L_g = 0)$. The robust estimation strategy improves the estimate of the algorithm in [3] by 11.0% and 9.5% at the 50th and 99th percentile.

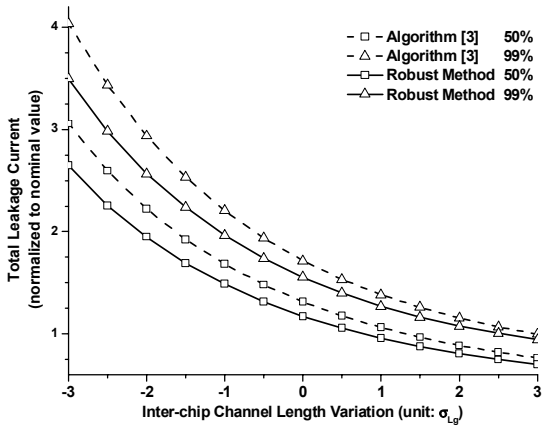

Figure 9. Equi-yield contours for the inter-chip *L* variation. The robust approach reduces the estimate of the analytical algorithm in [3] by 5.3-13.4% at the 99th percentile.

Table I. Normalized total leakage at the 99th percentile.

Inter-chip *L* variation (σ_{Lg})	-3	-2	-1	0	1	2	3
Algorithm in [3]	4.04	2.94	2.21	1.72	1.39	1.16	1.00
Robust modeling of V_{dd} and *T*	3.50	2.57	1.97	1.55	1.27	1.08	0.94
Reduction (%)	13.4	12.6	10.5	9.5	8.1	6.9	5.3

7. ACKNOWLEDGMENTS

This work was supported in part by GSRC, NSF, SRC, Sun, Intel, and the University of Texas at Austin.

8. REFERENCES

[1] S. Zhang, V. Wason, and K. Banerjee, "A probabilistic framework to estimate full-chip threshold leakage power distribution considering within-die and die-to-die P-T-V variations," in *Proc. of ISLPED*, 2004.

[2] H. Su *et al.*, "Full-chip leakage estimation considering power supply and temperature variations," in *Proc. of ISLPED*, 2003.

[3] R. Rao *et al.*, "Parametric yield estimation considering leakage variability," in *Proc. of DAC*, 2004.

[4] A. Srivastava *et al.*, "Accurate and efficient gate-level parametric yield estimation considering correlated variations in leakage power and performance," in *Proc. of DAC, 2005*.

[5] H. Chang and S. Sapatnekar, "Full-chip analysis of leakage power under process variations, including spatial correlations," in *Proc. of DAC*, 2005.

[6] W. -S. Wang, V. Kreinovich, and M. Orshansky, "Statistical timing based on incomplete probabilistic descriptions of parameter uncertainty," in *Proc. of DAC*, 2006.

[7] J. Ma and R. A. Rutenbar, "Interval-valued reduced order statistical interconnect modeling," in *Proc. of ICCAD*, 2004.

[8] D. Berleant and C. Goodman-Strauss, "Bounding the results of arithmetic operations on random variables of unknown dependency using intervals," *Reliable Computing*, vol. 4, no. 2, pp. 174-165, 1998.

[9] Dan Ernst *et al.*, "Razor: circuit-level correction of timing errors for low-power operation," *IEEE Micro*, vol. 24, no. 6, pp. 10-20, Nov. 2004.

[10] K. Skadron *et al.*, "Temperature-aware microarchitecture: modeling and implementation," *ACM Trans. Architecture and Code Optimization*, vol. 1, no. 1, pp. 94-125, Mar. 2004.

[11] S. Mukhopadhyay *et al.*, "Modeling and estimation of total leakage current in nano-scaled CMOS devices considering the effect of parameter variation," in *Proc. of ISLPED*, 2003.

[12] G. McLachlan and D. Peel, *Finite Mixture Models*, Wiley, 2000.

[13] W. Feller, *An Introduction to Probability Theory and Its Applications*, Wiley and Sons, 3rd Edition, 1968.

[14] S. Ferson *et al.*, "Constructing probability boxes and Dempster-Shafer structures," Sandia Report, 2002.

[15] H. Godwin, *Inequalities on Distribution Functions*, Hafner, 1964.

[16] H. Reagan *et al.*, "Equivalence of methods for uncertainty propagation of real-valued random variables," *Int. J. of Approximate Reasoning*, vol. 36, no. 1, pp. 1-30, Apr. 2004.

[17] D. Berleant and J. Zhang, "Using Pearson correlation to improve envelopes around the distributions of functions," *Reliable Computing*, vol. 10, no. 2, pp. 139-161, 2004.

[18] D. Lackey *et al.*, "Managing power and performance for System-on-Chip designs using voltage islands," in *Proc. of ICCAD*, 2002.

[19] A. Raychowdhury, S. Mukhopadhyay, and K. Roy, "Modeling and estimation of leakage in sub-90nm devices," in *Proc. of Int. Conf. on VLSI Design*, 2004.

[20] Predictive technology model. Available: http://www.eas.asu.edu/~ptm.

[21] Y. Cao *et al.*, "New paradigm of predictive MOSFET and interconnect modeling for early circuit design," in *Proc. of CICC*, 2000.