Statistical Timing Analysis with Two-sided Constraints

Khaled R. Heloue Farid N. Najm

Department of Electrical and Computer Engineering
University of Toronto, Toronto, Ontario, Canada
{khaled, najm}@eecg.toronto.edu

ABSTRACT

Based on a timing yield model, a statistical static timing analysis technique is proposed. This technique preserves existing methodology by selecting a "device file setting" that takes into account within-die statistical variations, and with which to run traditional static timing analysis in order to meet the desired yield. Using process-specific "generic paths" representing critical paths in a given process technology, our approach can be used early in the design process, most importantly during the pre-placement phase. Within-die variations are taken care of using a simple model that assumes positive correlation, which leads to upper and lower bounds on the timing yield. Our approach also handles both setup and hold timing constraints.

1. INTRODUCTION

The yield of an integrated circuit (IC) is a complex function of a number of factors related to both design and manufacturing. In this work, we will focus on the yield affected by circuit design. As part of circuit timing verification, one has to leave enough margin so that circuit delay variations do not affect yield too adversely. We will focus on this part of the overall yield problem, referred to as *timing yield* or *circuit-limited yield* [1, 2].

It is a known fact that process variations have an impact on circuit delay variations, and consequently can cause timing yield loss. Traditionally, process variations have been taken care of in various ways. ASICs are typically designed by making sure the chip passes the timing requirements at all *process corners*, including nominal, worst, and best cases of device behavior. A circuit is deemed to have *passed* the timing test if it meets the performance constraints for all "worst-case" files belonging to that process. If these settings are too pessimistic, then designers are forced to waste time and effort optimizing a circuit using design conditions that are too stringent. For microprocessors, it is typical to check circuit timing with *nominal* transistor files, and to specify some *timing margin* which should be left as slack in order to account for process variations.

There has been considerable discussion in the literature that the traditional methods of using process corners or using a timing margin are breaking down. For one thing, for ASICs, the number of corners is increasing, making it very expensive to explore all corners. Also, the corner based method can be too conservative and does not provide the user with any quantitative feedback on the robustness of the design [3]; it is a pass/fail approach. Furthermore, this traditional approach cannot handle *within-die* statistical variations [2]. On the other hand, for microprocessors, where nominal process files are used and a timing margin needs to be left as slack, there are no easy ways to decide what the margin should be, to account for within-die variations which have become important recently [1].

Statistical techniques offer an alternative approach; statistical transistor modeling techniques [3] have been used for quite some time. Recently, due to the increased importance of within-die variations, there has been an increased interest in tackling the timing yield problem by employing statistical techniques as part of the circuit timing analysis step [2, 1, 4, 5, 6]. The aim is to include statistical delay variations as an extension to traditional STA leading to statistical static timing analysis (SSTA).

In a number of cases [2, 6, 7, 8], it has been assumed that within-die variations are totally uncorrelated, an assumption which is not true in practice. It is usually hard to express the correlations between within-die parameter variations with a model built from process data. There are no published models, for instance, on how exactly the variations are correlated across the die as a function, say, of the distance between components. In [5], even though statistical within-die variations are *not* taken into account, a suggestion is made at the end as to how one may include them and take care of correlation by enforcing correlation between features that are in the same region of the layout. This theme was further developed, where use was made of principal components analysis (PCA) [9] or a quad-tree partitioning [10] to express a region-wise spatial correlation among within-die variations. Here too, it is not clear how one would identify these regions and how the model would be built from process data. Finally, since these methods depend on placement information, and extensive process data, these types of post-placement, design specific SSTA become final sign-off tools and are unusable during circuit design.

In fact, we envision that three types of SSTA may be useful in practice:

1. process-specific SSTA based on a generic path. This can be applied early in the design flow, to establish timing margins for generic paths in the candidate technology, even before circuit design has started, and to possibly optimize the devices or the circuit style to reduce these margins.
2. design-specific SSTA based on a given design in a given process. This can be applied pre-placement, during the circuit design stage. This would be perhaps the most heavily used type of SSTA.
3. post-placement design-specific SSTA, for final sign-off.

Of course, the level of accuracy achieved and the level of physical detail that is taken into account will vary among the different types of SSTA. For instance, wire parasitics variations have an effect on delay (depending on the strengths of the driver, etc.), but can only really be taken into account post-placement. In a pre-placement scenario, they may have to be ignored, or replaced by some safety factor.

In [11], a pre-placement, process specific SSTA is presented, in which within-die statistical correlations are captured with principal components analysis (PCA). Using Cauchy bounds, an approach is then developed to estimate a lower bound on the Max timing yield (i.e., based on setup constraints) of a large collection of generic paths. However, meeting the hold time constraints, reflected in the Min timing yield, was not covered in [11]. Also, the analysis is based on prior knowledge of the *order* of PCA, which may be hard to predict. In this work, we extend the approach of [11] in two important ways: 1) we can handle both setup and hold timing constraints and 2) our analysis is not dependent on knowing the order of the PCA for the within-die correlation model. The resulting approach is a pre-placement process-specific SSTA technique, also based on the "generic critical path" concept presented in [12, 11], to perform statistical timing analysis with two-sided constraints. Our approach can be used in the pre-placement phase, to establish timing margins for generic paths, even before circuit design has started. In our approach, we assume the within-die statistical variations to be arbitrarily *positively* correlated across the die, i.e., having a positive covariance, which is a reasonable assumption for many sources of variability. However, we do not require the availability of a correlation model. Instead, we will prove that by assuming the within-die systematic variations of path delays to be uncorrelated/totally correlated, we will produce lower/upper bounds on the Max and

Min timing yields. Using only extreme cases of correlation, this approach requires minimal knowledge of process data, and hence is applicable to the pre-layout phase, during circuit design and optimization.

Previously proposed techniques for statistical static timing analysis change the static timing flow so that one is propagating distributions of delay, instead of simply delay. In contrast, our approach does not propagate distributions. Instead, the result of our approach is a selection of a "device file" setting or a "virtual corner" with which to run traditional static timing analysis, which is somewhere within the extremes of device behavior. For example, while the "nominal" device file may call for a setting of $\Delta L = 0$ (for channel length variations) and the "worst-case" file may call for a setting of $\Delta L = +3\sigma_L$, our approach aims to predict the value "δ" such that if the setting of $\Delta L = \delta\sigma_L$ was used for all devices, and if the circuit timing is verified using traditional static timing analysis, then the circuit would give the desired timing yield. Another equivalent result is to be able to predict a timing margin τ which, if allowed, and if traditional STA with nominal files was used, the desired yield would be met. Hence, for a desired yield, we will work backwards to find the required timing margin τ or the required device file setting δ. As a result, our approach can be applied in the pre-layout phase. It preserves existing static timing methodology and only assumes the existence of statistical transistor models, which have been standard for some time.

2. PARAMETER MODEL

For a given circuit element or layout feature i, let $X(i)$ be a zero-mean Gaussian *random variable* (RV) that denotes the variation of a certain parameter of this element from its nominal (mean) value. Thus, for example, $X(i)$ may represent channel length variations of transistor i. Correlation between values of $X(i)$ at different locations on the die may be expressed by means of an autocorrelation function, but this is not a practical approach. Instead, it is standard practice [13] to express the correlation by first breaking up the variations into *die-to-die* and *within-die* components, as follows:

$$X(i) = X_{dd} + X_{wd}(i) \tag{1}$$

The die-to-die component X_{dd} is an *independent*[1] zero-mean Gaussian RV that takes the same value for all instances of this element on a given die, irrespective of location. The within-die component $X_{wd}(i)$ is a zero-mean Gaussian which can take different values for different instances of that element on the same die. This leads to the following relationship between the variances:

$$\sigma^2(i) = \sigma^2_{dd} + \sigma^2_{wd}(i) \tag{2}$$

Then, the within-die component is further broken down into two components, a *systematic* component and a "random" component:

$$X_{wd}(i) = X_{wds}(i) + X_{wdr}(i) \tag{3}$$

where, for each i, the random component $X_{wdr}(i)$ is an *independent* zero-mean Gaussian. A similar relationship follows for the variances:

$$\sigma^2_{wd}(i) = \sigma^2_{wds}(i) + \sigma^2_{wdr}(i) \tag{4}$$

We can write $X_{wds}(i)$ in the following way:

$$X_{wds}(i) = \sigma_{wds}(i)\, Z_{wds}(i) \tag{5}$$

where $Z_{wds}(i)$ are correlated standard normal RVs (mean 0, variance 1). Hence, our model for parameter variation $X(i)$ consists of an independent zero mean die-to-die component X_{dd} with variance σ^2_{dd}, a correlated systematic within-die component $X_{wds}(i)$ with variance $\sigma^2_{wds}(i)$, and an independent random within-die component $X_{wdr}(i)$ with variance $\sigma^2_{wds}(i)$.

[1]Throughout this paper, whenever an individual RV is described as "independent", this means that it is independent of all other RVs under consideration.

In previous work, PCA and dependence on global sources of variations were used to model correlation between the systematic components of the within-die variations. In this work however, we will not require the use of a specific correlation model. Instead, we express the systematic components as positively correlated RVs, and we show that our timing yield bounds hold for any correlation model, and are the same for any order of the PCA expansion.

3. PARAMETRIC YIELD MODEL

With the random parameter model given above, we now define the *parametric Max yield* and the *parametric Min yield* for parameter X as:

$$Y_{max}(x) = \mathcal{P}\{X(i) \le x,\ i = 1, 2, \ldots, n\} \tag{6}$$

$$Y_{min}(x) = \mathcal{P}\{X(i) > -x,\ i = 1, 2, \ldots, n\} \tag{7}$$

where n is the number of instances of this parameter on chip. Here, $X(i)$ is a *generic* parameter that may represent transistor channel length variations, threshold voltage variations, etc. In fact, $X(i)$ is *any* statistical quantity on chip that may be characterized by the parameter model introduced in section 2. When $X(i)$ is a simple parameter, such as channel length, then parametric Max yield is the probability that *all* device lengths on the die vary by less than some threshold x, and parametric Min yield is the probability that all device length variations are greater than some other threshold $-x$. Note that x is generally positive since in the Max yield case, the threshold is usually set greater than the mean of X (i.e zero), while in the Min yield case it is usually set to be smaller. We will later show how path delay can itself be viewed as a *parameter* with its own triplet of variances (σ^2_{dd}, σ^2_{wds}, σ^2_{wdr}) that we will relate to the underlying transistor parameter variances. This will allow us to express timing yield based on a parametric yield model. Thus, the material in this section, although focused on parametric yield, will actually be directly useful for computing timing yield.

It is noteworthy that both Max and Min yields are of equal importance. Chips fail either because the maximum circuit delay is greater than a performance constraint (the setup constraint), or because the minimum circuit delay is smaller than another constraint (the hold constraint).

3.1 Slepian's Inequality

In this section we will find bounds on the two parametric yields that were defined earlier. Using the following theorem from multivariate normal probability [14], we will prove that these bounds are a direct consequence of extreme cases of correlation between the within-die systematic components across the die.

Slepian's Inequalities:
Let $V = (\{V(i)\},\ i = 1, \ldots, n)$ and $W = (\{W(i)\},\ i = 1, \ldots, n)$ be two random vectors of size n, both multi-normally distributed with zero mean and covariance matrices $\Sigma = \{\sigma_{ij}\}$ and $\Gamma = \{\gamma_{ij}\}$, respectively. Also let $\sigma_{ii} = \gamma_{ii}$ for $i = 1, \ldots, n$. If $\sigma_{ij} \ge \gamma_{ij}$ for all $i \ne j$, in symbols $\Sigma \ge \Gamma$, then:

$$\mathcal{P}\{V(i) \le a_i,\ \forall i\} \ge \mathcal{P}\{W(i) \le a_i,\ \forall i\} \tag{8}$$

and

$$\mathcal{P}\{V(i) > a_i,\ \forall i\} \ge \mathcal{P}\{W(i) > a_i,\ \forall i\} \tag{9}$$

Note that $\sigma_{ii} = \sigma_i^2$ is the variance of $V(i)$, and σ_{ij} is the covariance of $V(i)$ and $V(j)$.

The above result is quite interesting. It states that, given two random vectors V and W having the same variances (i.e., $\sigma_{ii} = \gamma_{ii}$), if one of them is more correlated than the other (i.e., $\sigma_{ij} \ge \gamma_{ij}$ for all $i \ne j$) then (8) and (9) hold.

A direct consequence of Slepian's inequalities is to be able to find upper and lower bounds on both Max and Min parametric yields. As presented in section 2, the parameter vector X is the following:

$$X(i) = X_{dd} + X_{wds}(i) + X_{wdr}(i) \tag{10}$$

where $X_{wds}(i)$'s are arbitrarily correlated across the die. To overcome the lack of information about within-die systematic correlation, which is typically hard to get, we will introduce two RVs

that represent extreme cases of within-die systematic correlation: the case of independence, and the case of total correlation.

Let $X^{(0)}$ be a random vector whose elements have the same marginal distributions as those of X, but with the property that its within-die systematic components are independent. Also, let $X^{(1)}$ be a random vector whose elements have the same marginal distributions as those of X, but with the property that its within-die systematic components are positively totally correlated. In other words, X, $X^{(0)}$, and $X^{(1)}$ have the same individual variances, but only differ in the extent of correlation of the within-die systematic components, i.e., they differ in their covariances.

Now, suppose that Σ, $\Sigma^{(0)}$ and $\Sigma^{(1)}$ are the covariance matrices of parameter vectors X, $X^{(0)}$, and $X^{(1)}$ respectively. These matrices have the same diagonal elements (representing variances), but differ by their off-diagonal elements (representing covariances). Remember that X is a zero-mean random vector, and its covariance matrix $\Sigma = \{\sigma_{ij}\}$ will be simplified to the following, for $i \neq j$:

$$\begin{aligned} \sigma_{ij} &= E\left[X(i) \cdot X(j)\right] \\ &= \sigma_{dd}^2 + E\left[X_{wds}(i) \cdot X_{wds}(j)\right] \end{aligned} \tag{11}$$

where σ_{ij} is the off-diagonal element of the covariance matrix Σ. The second part of the above equation is the covariance $\Sigma_{wds} = \{\sigma_{wds_{ij}}\}$ of the within-die systematic component. Therefore we can write:

$$\sigma_{ij} = \sigma_{dd}^2 + \sigma_{wds_{ij}} \tag{12}$$

Notice here that the covariance of $X^{(0)}$ and $X^{(1)}$ will have the same form as the above equation, and only differ by the value of $\sigma_{wds_{ij}}$: For $X^{(0)}$, $\sigma_{wds_{ij}} = 0$ since the within-die systematic components are independent (correlation coefficient equals to 0). For $X^{(1)}$, $\sigma_{wds_{ij}}$ is maximum, since maximum covariance occurs when the systematic RVs are positively totally correlated (correlation coefficient equals to +1).

If we assume that the within-die systematic components are positively correlated (i.e., their covariances are always positive $\sigma_{wds_{ij}} \geq 0$), then the correlation coefficients are between the two extremes of 0 and 1, so that:

$$\Sigma^{(0)} \leq \Sigma \leq \Sigma^{(1)} \tag{13}$$

The assumption of positive correlation is practical for many sources of variability. A physical variation that slows down a transistor, a gate, or a path, is likely to have the same effect on another that lies nearby. If the other device/gate/path is far away, then it would probably be independent. Given this assumption, then coupling (13) with Slepian's inequalities, we can write:

$$\mathcal{P}\{X^{(0)}(i) \leq x,\ \forall i\} \leq Y_{max}(x) \leq \mathcal{P}\{X^{(1)}(i) \leq x,\ \forall i\} \tag{14}$$

$$\mathcal{P}\{X^{(0)}(i) > -x,\ \forall i\} \leq Y_{min}(x) \leq \mathcal{P}\{X^{(1)}(i) > -x,\ \forall i\} \tag{15}$$

where $Y_{max}(x)$ and $Y_{min}(x)$ are the parametric Max and Min yield defined earlier. The above equations are fundamental, because they capture the lower and upper bounds on the Max and Min yields when the within-die correlation is unknown or uncertain. In the following sections, we will derive expressions for the lower bounds and upper bounds given in (14) and (15).

3.2 Lower bounds

Consistent with the previous section, we will now find expressions for the Max and Min yields lower bounds, by starting with an assumption that $X_{wds}(i)$ are independent. Let $Y_{max}^{(lb)}(x)$ and $Y_{min}^{(lb)}(x)$ be the Max and Min yield lower bounds respectively. For the lower bound analysis, the within-die systematic and random components $X_{wds}(i)$ and $X_{wdr}(i)$ are both independent RVs. Therefore we can replace them both by an independent within-die component $X_{wd}(i) = X_{wds}(i) + X_{wdr}(i)$. The following are the yield lower bound expressions:

$$Y_{max}^{(lb)}(x) = \mathcal{P}\left\{X_{dd} + X_{wd}(i) \leq x,\ \forall i\right\} \tag{16}$$

$$Y_{min}^{(lb)}(x) = \mathcal{P}\left\{X_{dd} + X_{wd}(i) > -x,\ \forall i\right\} \tag{17}$$

Since X_{dd} is an independent zero-mean Gaussian with variance σ_{dd}^2, then $Z_0 = X_{dd}/\sigma_{dd}$ is an independent standard normal RV (mean 0, variance 1), and the expressions for the yield lower bounds can be expanded as:

$$Y_{max}^{(lb)}(x) = \mathcal{P}\{\sigma_{dd}Z_0 + X_{wd}(i) \leq x,\ \forall i\} \tag{18}$$

$$Y_{min}^{(lb)}(x) = \mathcal{P}\left\{\sigma_{dd}Z_0 + X_{wd}(i) > -x,\ \forall i\right\} \tag{19}$$

We now recall a result from basic probability theory that will be used repeatedly in the paper. Let $\mathcal{A}$ be an arbitrary event, and X be an RV with a probability density function (pdf) $f(x)$. Then (see [15], pg. 85) we have:

$$\mathcal{P}\{\mathcal{A}\} = \int_{-\infty}^{+\infty} \mathcal{P}\{\mathcal{A} \mid X = x\} f(x) dx \tag{20}$$

This result is an extension to the continuous case of the simple fact that $\mathcal{P}\{\mathcal{A}\} = \mathcal{P}\{\mathcal{A} \mid \mathcal{B}\} \cdot \mathcal{P}\{\mathcal{B}\} + \mathcal{P}\{\mathcal{A} \mid \overline{\mathcal{B}}\} \cdot \mathcal{P}\{\overline{\mathcal{B}}\}$, where $\mathcal{B}$ is another event. Applying (20) to (18) and (19), and denoting by $\phi(\cdot)$ the pdf of the standard normal distribution, gives:

$$Y_{max}^{(lb)}(x) = \int_{-\infty}^{+\infty} \mathcal{P}\{X_{wd}(i) \leq x - \sigma_{dd}z,\ \forall i\}\phi(z)dz \tag{21}$$

$$Y_{min}^{(lb)}(x) = \int_{-\infty}^{+\infty} \mathcal{P}\{X_{wd}(i) > -x - \sigma_{dd}z,\ \forall i\}\phi(z)dz \tag{22}$$

Since $X_{wd}(i)$ are independent, then we can express their joint probability as a product:

$$Y_{max}^{(lb)}(x) = \int_{-\infty}^{+\infty} \prod_{i=1}^{n} \mathcal{P}\{X_{wd}(i) \leq x - \sigma_{dd}z\}\phi(z)dz \tag{23}$$

$$Y_{min}^{(lb)}(x) = \int_{-\infty}^{+\infty} \prod_{i=1}^{n} \mathcal{P}\{X_{wd}(i) > -x - \sigma_{dd}z\}\phi(z)dz \tag{24}$$

Now replacing the integrals by the mean or expected value operator $E[\cdot]$ and using the fact that for a normal RV W with zero mean, $\mathcal{P}\{W > -w\} = \mathcal{P}\{W \leq w\}$, we get:

$$Y_{max}^{(lb)}(x) = E\left[\prod_{i=1}^{n} \Phi\left(\frac{x - \sigma_{dd}Z_0}{\sigma_{wd}(i)}\right)\right] \tag{25}$$

$$Y_{min}^{(lb)}(x) = E\left[\prod_{i=1}^{n} \Phi\left(\frac{x + \sigma_{dd}Z_0}{\sigma_{wd}(i)}\right)\right] \tag{26}$$

where $\sigma_{wd}^2(i) = \sigma_{wds}^2(i) + \sigma_{wdr}^2(i)$ is the variance of $X_{wd}(i)$, and $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard normal.

Equations (25) and (26) represent the parametric Max and Min yield lower bounds. Note the dependence on the number of parameter instances n. Later in this paper, we will show that as n goes to infinity, the above equations will take other forms which are independent of n.

3.3 Upper bound

Similarly, we will now find expressions for upper bounds on the parametric Max and Min yields. Recall from section 3.1 that when $X_{wds}(i)$ are positively totally correlated, then one gets an upper bound on the parametric yields. The following equations represent upper bounds on the parametric yield, where $X_{wds}(i)$ are totally correlated.

$$Y_{max}^{(ub)}(x) = \mathcal{P}\left\{X_{wdr}(i) + X_{wds}(i) + \sigma_{dd}Z_0 \leq x,\ \forall i\right\} \tag{27}$$

$$Y_{min}^{(ub)}(x) = \mathcal{P}\left\{X_{wdr}(i) + X_{wds}(i) + \sigma_{dd}Z_0 > -x,\ \forall i\right\} \tag{28}$$

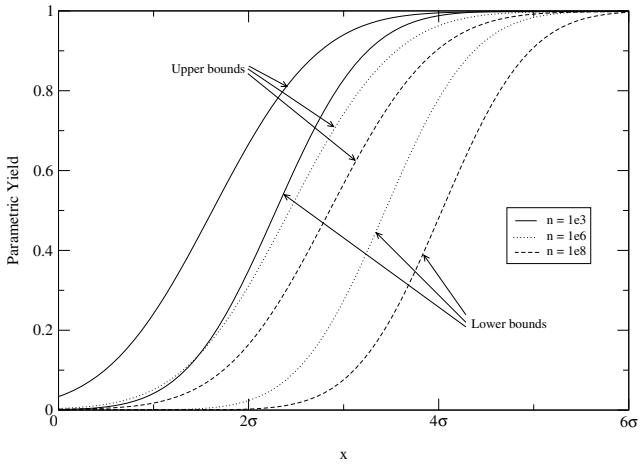

Figure 1: Parametric Max yield bounds

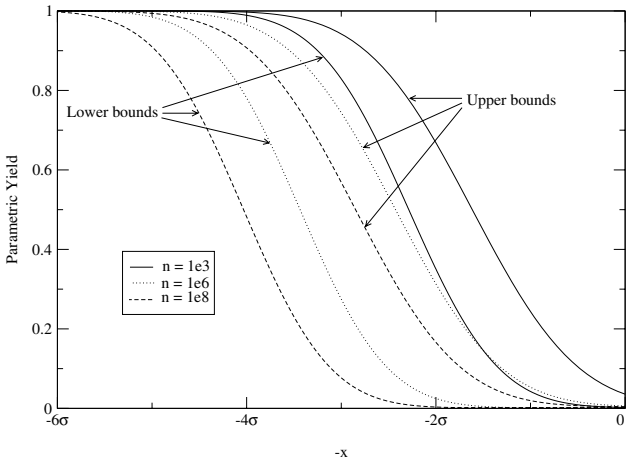

Figure 2: Parametric Min yield bounds

Total correlation of the within-die systematic components means that there is a unique random variable that models all the systematic variations across the die. Hence the model for $X_{wds}(i)$ will be the following:

$$X_{wds}(i) = \sigma_{wds}(i) Z_1 \tag{29}$$

where Z_1 is an independent standard normal. Making use of (20) twice on variables Z_0 and Z_1, and noting that $X_{wdr}(i)$ are independent with variance $\sigma^2_{wdr}(i)$, we get the following expressions:

$$Y^{(ub)}_{max}(x) = E\left[\prod_{i=1}^{n} \Phi\left(\frac{x - \sigma_{dd} Z_0 - \sigma_{wds}(i) Z_1}{\sigma_{wdr}(i)}\right)\right] \tag{30}$$

$$Y^{(ub)}_{min}(x) = E\left[\prod_{i=1}^{n} \Phi\left(\frac{x + \sigma_{dd} Z_0 + \sigma_{wds}(i) Z_1}{\sigma_{wdr}(i)}\right)\right] \tag{31}$$

The same analysis used in the lower bound case is applied here to get the above equations. For sake of conciseness, several steps were omitted.

As a result of sections 3.2 and 3.3, we have an upper and lower bound on each of the parametric Max and Min yields. With a simple change of variables, as will be illustrated in the next section, these bounds can be computed by numerical integration. If a Max or Min yield of, say, better than 90% is desired, then one can set the lower bounds $Y^{(lb)}_{max}(x)$ or $Y^{(lb)}_{min}(x)$ to 0.9 and work backwards to get the value of the threshold x. Upper and lower bounds are the results of extreme cases of within-die systematic correlations. Therefore when combined together, they become very useful in estimating the yield when the correlations are uncertain or unknown.

3.4 Illustration

As an illustration, we will assume all variances to be the same across the die. This means $\sigma^2_{wds}(i) = \sigma^2_{wds}$ and $\sigma^2_{wdr}(i) = \sigma^2_{wdr}$. Then, we can rewrite equations (25), (26), (30), and (31) in the following way:

$$Y^{(lb)}_{max}(x) = E\left[\Phi^n\left(\frac{x - \sigma_{dd} Z_0}{\sigma_{wd}}\right)\right] \tag{32}$$

$$Y^{(lb)}_{min}(x) = E\left[\Phi^n\left(\frac{x + \sigma_{dd} Z_0}{\sigma_{wd}}\right)\right] \tag{33}$$

$$Y^{(ub)}_{max}(x) = E\left[\Phi^n\left(\frac{x - \sigma_{dd} Z_0 - \sigma_{wds} Z_1}{\sigma_{wdr}}\right)\right] \tag{34}$$

$$Y^{(ub)}_{min}(x) = E\left[\Phi^n\left(\frac{x + \sigma_{dd} Z_0 + \sigma_{wds} Z_1}{\sigma_{wdr}}\right)\right] \tag{35}$$

where σ_{wd} was defined earlier to be $\sqrt{\sigma^2_{wds} + \sigma^2_{wdr}}$.

In order to compute these equations, we use the definition of the expected value operator (as an integral) and with a change of variables of $u = \Phi(z_0)$ and $v = \Phi(z_1)$, we arrive at:

$$Y^{(lb)}_{max}(x) = \int_0^1 \Phi^n\left(\frac{x - \sigma_{dd}\Phi^{-1}(u)}{\sigma_{wd}}\right) du dv \tag{36}$$

$$Y^{(lb)}_{min}(x) = \int_0^1 \Phi^n\left(\frac{x + \sigma_{dd}\Phi^{-1}(u)}{\sigma_{wd}}\right) du dv$$

$$Y^{(ub)}_{max}(x) = \int_0^1 \int_0^1 \Phi^n\left(\frac{x - \sigma_{dd}\Phi^{-1}(u) - \sigma_{wds}\Phi^{-1}(v)}{\sigma_{wdr}}\right) du dv$$

$$Y^{(ub)}_{min}(x) = \int_0^1 \int_0^1 \Phi^n\left(\frac{x + \sigma_{dd}\Phi^{-1}(u) + \sigma_{wds}\Phi^{-1}(v)}{\sigma_{wdr}}\right) du dv$$

Plots of the bounds on the parametric Max and Min yields for different values of n are shown in Fig. 1 and Fig. 2. In these plots, we have assumed $\sigma^2_{dd} = 0.5\sigma^2$, and $\sigma^2_{wds} = \sigma^2_{wdr} = 0.25\sigma^2$, where σ^2 is the total variance. Notice that yield decreases for larger n, as expected. Also note that as x increases ($-x$ decreases), it is more likely that the max of $X(i)$ is less than x, and the min of $X(i)$ is greater than $-x$.

3.5 Bounded Variations

Notice that, in the above expressions for yield, the yield decreases for larger n. One would somewhat expect this, but it is surprising to note that the yield approaches zero as n goes to infinity, for any combination of values of the three variances. This may also be seen in the above plots in Fig. 1. This is somewhat non-physical, and arises due to the fact that we have assumed that the distribution of $X_{wdr}(i)$ is normal; recall that the normal distribution extends to $\pm\infty$ in both directions. In reality, one would expect process variations to be bounded by some upper and lower bounds. If a device somewhere deviates by large amounts, like 6σ or 7σ, then chances are there is a serious problem with that die, and that it would be lost due to other reasons, other than timing yield that is. Therefore, it is a good idea to limit the spread of the cdf of $X_{wdr}(i)$ to some multiple of σ in order to avoid these non-physical effects at large n. In this section, therefore, we will use a *truncated normal* distribution for $X_{wdr}(i)$. For clarity of presentation, we will restrict the analysis to the illustrative special case introduced in section 3.4 where all variance are the same across the die. The analysis can be extended to the general case. Suppose, therefore, that $X_{wdr}(i)$ is bounded by $\pm k\sigma$, and let $\Phi_t(x)$ represent the cdf of the truncated standard normal, which is 0 for $x \leq -k$ and 1 for $x \geq k$.

We can plug $\Phi_t(\cdot)$ instead of $\Phi(\cdot)$ (for $X_{wdr}(i)$) into the above equations and plot the resulting yield integrals, as shown in Figs. 3 and 4. In this case the yield loss at higher n values is limited so that the 1e6 and 1e8 plots in each group are indistinguishable. This is to be expected, because the "tail" of the distribution has

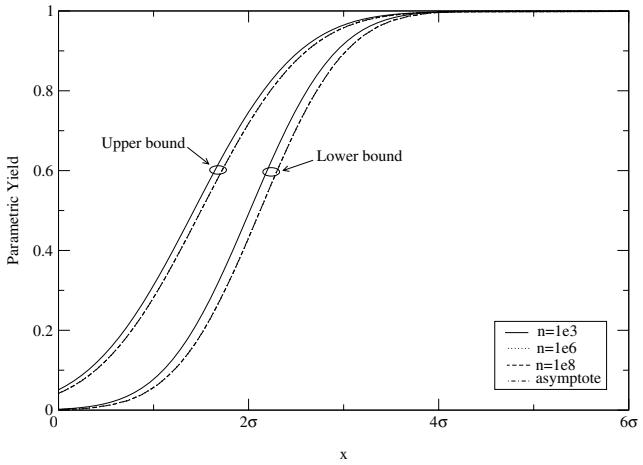

Figure 3: Parametric Max yield bounds for $k = 3$

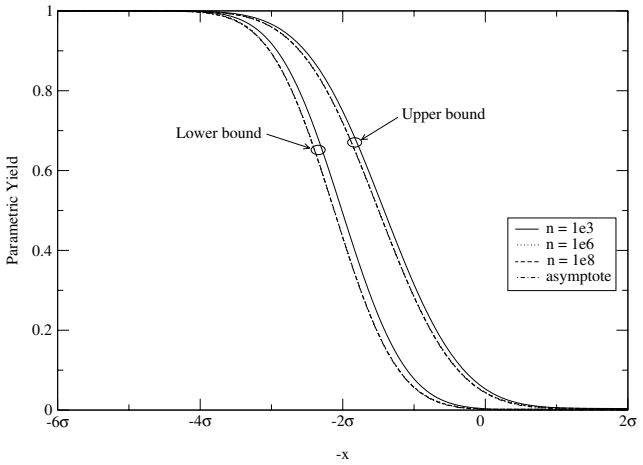

Figure 4: Parametric Min yield bounds for $k = 3$

been cut off, and it is primarily the tail that causes the yield loss at very large n.

When working with a truncated normal, it is noteworthy that we can derive asymptotes on the parametric yield curves that are independent of n. This means that as n tends to infinity, the parametric yield curves will be equal to these asymptotes. The derivations are not shown, for brevity, but lead to the following results, where $Y_0(x) = \lim_{n\to\infty} Y(x)$:

$$Y_{0_{max}}^{(lb)}(x) = \Phi\left(\frac{x - k\sigma_{wd}}{\sigma_{dd}}\right) \quad (37)$$

$$Y_{0_{min}}^{(lb)}(x) = \Phi\left(\frac{x - k\sigma_{wd}}{\sigma_{dd}}\right) \quad (38)$$

$$Y_{0_{max}}^{(ub)}(x) = E\left[\Phi\left(\frac{x - \sigma_{dd}Z_0 - k\sigma_{wdr}}{\sigma_{wds}}\right)\right] \quad (39)$$

$$Y_{0_{min}}^{(ub)}(x) = E\left[\Phi\left(\frac{x + \sigma_{dd}Z_0 - k\sigma_{wdr}}{\sigma_{wds}}\right)\right] \quad (40)$$

where, as before, $Z_0 = X_{dd}/\sigma_{dd}$ is an independent standard normal RV. Equations (37) and (38), especially when applied to timing yield as we will do later in the paper, exhibit similar observations as was made in [12]. Namely, the within-die variations determine the mean of the yield, while the die-to-die variations determine the spread of the yield. Fig. 3 and Fig. 4 show the plots of these asymptotes for both parametric Max and Min yields. These asymptotes are very tight, and indistinguishable on the plot from the 1e6 and 1e8 curves.

4. TIMING YIELD MODEL

In this section, we will show that path delay can be handled as a parameter. Hence, we will use the parametric yield analysis to get bounds on both timing Max and Min yields. Using the same approach as [11], we will start from the transistor level with channel length and threshold voltage variations, $L(i)$ and $V(i)$ with variances $\sigma_L^2(i)$ and $\sigma_V^2(i)$ respectively, and move to the gate level assuming that gate delay variation is a linear combination of transistor variations, mainly $D(i) = \alpha L(i) + \beta V(i)$, with variance $\sigma_D^2(i) = \alpha^2\sigma_L^2(i) + \beta^2\sigma_V^2(i)$. All these variances can be decomposed into die-to-die and within-die components, as presented earlier.

4.1 Bounds on Sum of RVs

In this section, we will find bounds on the distribution of the sum of N normal RVs $X(i)$ which will turn out to be useful for the chip timing yield analysis. Assuming that these RVs are expressed in the same way as the parameter model defined in section 2, we will prove that a lower bound on such distribution is generated when the within-die systematic variations of $X(i)$'s are assumed totally correlated (correlation coefficients $\rho_{wds} = +1$), and an upper bound is generated when systematic variations of $X(i)$'s are independent ($\rho_{wds} = 0$).

Let $S = \sum_{i=1}^{N} X(i)$, then S is a zero-mean normal RV with some variance, call it σ_s^2, and we can write:

$$\mathcal{P}\{S \le x\} = \Phi(x/\sigma_s) \quad (41)$$

where $x \ge 0$. Since $\Phi(\cdot)$ is increasing, then $\mathcal{P}\{S \le x\}$ is decreasing in σ_s. Looking at the variance of S:

$$\sigma_s^2 = E[S^2] = \sum_{i=1}^{N} \sigma_{X_i}^2 + 2\sum_{i\ne j} E[X(i)\cdot X(j)] \quad (42)$$

$$= \sum_{i=1}^{N} \sigma_{X_i}^2 + 2\sum_{i\ne j} \sigma_{ij} \quad (43)$$

where σ_{ij} is the covariance of $X(i)$ and $X(j)$ found in (12). It is easy to show that, assuming the within-die systematic variations to be totally correlated (else independent) would set the within-die systematic covariances $\sigma_{wds_{ij}}$ (see (12)) to a maximum (else a minimum). Given (12), this makes the covariance σ_{ij} maximum (else minimum), and given (43), this also makes the variance σ_s^2 of the sum S maximum (else minimum), which finally produces a lower (else upper) bound on (41). The same reasoning is used to produce lower and upper bounds on $\mathcal{P}\{S > -x\}$, where x is positive, since $\mathcal{P}\{S > -x\} = \mathcal{P}\{S \le x\}$ in the case of a zero-mean normal RV.

In practice, one is quite interested in the yield *lower bound* expressions. Table 1 summarizes the conditions required to get a lower bound on the Max (Setup time) yield and the Min (Hold time) yield. It summarizes the results of this section along with section 3.1, to show that: In case of path delay, where gate delays are the "parameters", in order to produce a lower bound on yield, one must set the within-die systematic variations of gate delays to be totally correlated ($\rho_{wds} = +1$). Looking at block delay, where delay is the maximum/minimum of a large number of paths, with path delays as the "parameters", and in order to produce a lower bound on both Setup (Max) and Hold (Min) yields, one must set the within-die systematic variations of path delays to be independent ($\rho_{wds} = 0$).

4.2 Path Delay

Consider a generic path of N logic stages (a stage is a logic gate and the interconnect at its output). We will only focus on gate delays, and only on transistor L and V_t variations. The methodology can be easily applied when more device parameters are of interest, or when interconnect parameter variations are to be included as well. Let $D_N(j)$ denote the deviation of the delay

Table 1: Required settings for a yield lower bound

Yield Lower bound	Setup or Max Yield	Hold or Min Yield
Path delays	Sum over gates $\rho_{wds} = 1$	Sum over gates $\rho_{wds} = 1$
Block delay	Max over paths $\rho_{wds} = 0$	Min over path $\rho_{wds} = 0$

of path j from its mean (nominal) value:

$$D_N(j) = \sum_{i=1}^{N} D(i) = ND_{dd} + \sum_{i=1}^{N} D_{wds}(i) + \sum_{i=1}^{N} D_{wdr}(i) \quad (44)$$

Since path delay is the sum of a number of RVs (gate delays), then the results of section 4.1 become immediately useful. Thus, if we are computing the timing yield lower bounds, we will assume that (the systematic components of) gate delays on a given path are totally correlated. Else, if we are computing upper bounds, we will assume that (the systematic components of) gate delays are independent. Then, the delay of path j has a variance of:

$$\sigma^2_{D_N}(j) = \sigma^2_{dd,D_N} + \sigma^2_{wds,D_N}(j) + \sigma^2_{wdr,D_N}(j) \quad (45)$$

where:

$$\begin{aligned} \sigma^2_{dd,D_N} &= N^2\sigma^2_{dd,D} \\ \sigma^2_{wds,D_N}(j) &= N^2\sigma^2_{wds,D}(j) \text{ for lower bound} \\ \sigma^2_{wds,D_N}(j) &= N\hat{\sigma}^2_{wds,D}(j) \text{ for upper bound} \\ \sigma^2_{wdr,D_N}(j) &= \sum_{i=1}^{N}\sigma^2_{wdr,D}(i) = N\hat{\sigma}^2_{wdr,D}(j) \end{aligned}$$

and where $\hat{\sigma}^2_{wdr,D}(j)$ and $\hat{\sigma}^2_{wds,D}(j)$ are the average value of $\sigma^2_{wdr,D}(i)$ and $\sigma^2_{wds,D}(i)$ over all N gates on this path.

With this, we have a full statistical model of path delay, so that we can treat it as a "parameter" and we can talk about its Max and Min yields, as was done for the generic parameter $X(i)$ in section 3.

4.3 Timing Yield

The timing failure of an integrated circuit depends on a number of factors, including max delay violations, min delay violations, clock skew violations, etc. In this work, we focus on both max and min delay constraints and consider a circuit to "pass" the timing test if its longest (max critical) path delays are below some threshold, and its shortest (min critical) path delays are above some threshold. We let N_1 and N_2 be the number of stages (gates) on a path that would be representative of these long critical paths and short critical paths, respectively. We also consider that the chip contains a (typically large) number of *disjoint* (non-intersecting) critical paths of N_1 and N_2 stages (gates) each, so that our expressions for the *chip timing yields* becomes:

$$Y_{max}(\tau_1) = \mathcal{P}\left\{D_{N_1}(j) \le \tau_1,\ \forall j\right\} \ge Y_{0_{max}}^{(lb)}(\tau_1) \quad (46)$$

$$Y_{min}(\tau_2) = \mathcal{P}\left\{D_{N_2}(j) > -\tau_2,\ \forall j\right\} \ge Y_{0_{min}}^{(lb)}(\tau_2) \quad (47)$$

where $Y_{0_{max}}^{(lb)}(\cdot)$ and $Y_{0_{min}}^{(lb)}(\cdot)$ are the lower bound expressions for the yields found in section 3.5, with path delay considered as the "parameter". Since the paths being considered are disjoint, then any correlations between their delays is due only to correlations in the process variations, and not to the sharing of circuit component. If $\mathcal{Y}$ is the desired Max yield, then the techniques of section 3 effectively provide the inverse function to compute τ_1 for any desired $\mathcal{Y}$:

$$\tau_1 = {Y_{0_{max}}^{(lb)}}^{-1}(\mathcal{Y}) \quad (48)$$

This τ_1 is the timing margin of an N_1-gate path, for the desired specified Max yield $\mathcal{Y}$. Therefore, in order to get the desired yield, the circuit should be designed to "pass" the timing constraints when $D_{N1}(j) = \tau_1$, for all j. Therefore, we set $D(i) = \tau_1/N_1$. To get the transistor setting that will give the desired $\mathcal{Y}$ given the timing margin τ_1, we will use the same approach found in [11], from which the final equation for the transistor file setting is:

$$\delta = \frac{{Y_{0_{max}}^{(lb)}}^{-1}(\mathcal{Y})/N_1}{\alpha\sigma_L(i) + \beta\sigma_V(i)} \quad (49)$$

This δ effectively defines the "worst-case file" for which the circuit should be tested for timing constraint violations, so as to guarantee that the timing Max yield is at least $\mathcal{Y}$.

The same analysis can be performed for the timing Min yield, by replacing τ_1 and N_1 by τ_2 and N_2. Therefore, both max and min delay constraints can be met, by setting the Max and Min timing yields to the desired yield $\mathcal{Y}$, and get the corresponding transistor file setting.

5. APPLICATION TO TIMING ANALYSIS

We will now illustrate how the above timing yield model allows us to choose a setting δ for the transistor parameters so that a desired yield is achieved if the circuit passes traditional static timing analysis with that δ setting. We will use the timing Max yield as an illustration, but the same analysis can be done for the timing Min yield. Due to space limitations, this section will be very brief; it follows the similar development in [11]. For clarity, we will assume the same variance ratios as before, i.e., at the transistor level:

$$\sigma^2_{dd,L} = \frac{\sigma^2_L}{2}, \qquad \sigma^2_{wds,L}(i) = \sigma^2_{wdr,L}(i) = \frac{\sigma^2_L}{4}\ ,\ \forall i \quad (50)$$

$$\sigma^2_{dd,V} = \frac{\sigma^2_V}{2}, \qquad \sigma^2_{wds,V}(i) = \sigma^2_{wdr,V}(i) = \frac{\sigma^2_V}{4}\ ,\ \forall i \quad (51)$$

At the gate level, this leads to:

$$\begin{aligned} \sigma^2_{dd,D} &= \frac{1}{2}\left(\alpha^2\sigma^2_L + \beta^2\sigma^2_V\right) \\ \sigma^2_{wds,D}(x_i, y_i) &= \frac{1}{4}\left(\alpha^2\sigma^2_L + \beta^2\sigma^2_V\right) \\ \sigma^2_{wdr,D}(i) &= \frac{1}{4}\left(\alpha^2\sigma^2_L + \beta^2\sigma^2_V\right) \end{aligned} \quad (52)$$

Therefore, $\sigma^2_D = \left(\alpha^2\sigma^2_L + \beta^2\sigma^2_V\right)$ and, at the path level, we have:

$$\begin{aligned} \sigma^2_{dd,D_{N_1}} &= \frac{N_1^2}{2}\sigma^2_D \\ \sigma^2_{wds,D_{N_1}}(j) &= \frac{N_1^2}{4}\sigma^2_D \text{ for lower bound} \\ &= \frac{N_1}{4}\sigma^2_D \text{ for upper bound} \\ \sigma^2_{wdr,D_{N_1}}(j) &= \frac{N_1}{4}\sigma^2_D \end{aligned}$$

The last two equations are notable for the absence of the square factor in N_1, since they result from averaging of variations due to independence. Let $\sigma^2_{D_{N_1}}$ be the total path variance, then we can compute the timing yield at different multiples of $\sigma_{D_{N_1}}$, as shown in Fig. 5. This figure shows upper and lower bounds on the timing Max yield for $k = 3$ and $N_1 = 9$. One can use this type of figure as follows: If we want 95% Max yield (i.e., $\mathcal{Y} = 0.95$), then we will look at the plot of the **lower bound** and find that we need $\tau_1 \approx 3.1\sigma_{D_{N_1}}$, which, using (49), leads to:

$$\delta = \left(\frac{3.1\sigma_{D_{N_1}}/N_1}{\alpha\sigma_L + \beta\sigma_V}\right) \quad (53)$$

Now using the fact that $\sigma_{D_{N_1}} = N_1\sigma_D\sqrt{0.75 + 0.25N_1^{-1}}$ resulting from adding the path variances in the lower bound case, and

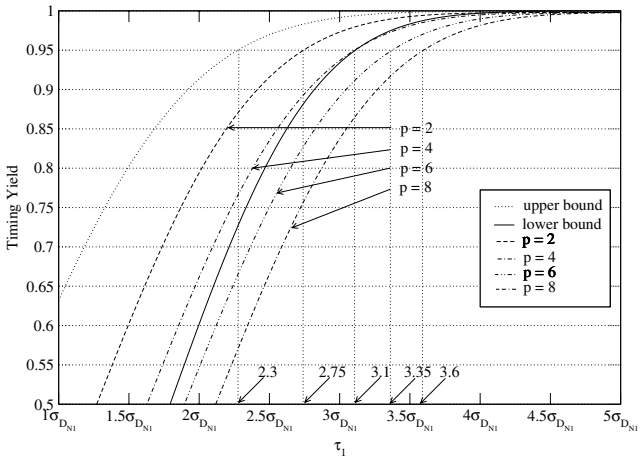

Figure 5: Timing Max yield bounds, for $k = 3$ and $N_1 = 9$, compared with Cauchy lower bounds

$N_1 = 9$, we get:

$$\delta = 2.73 \frac{\sqrt{\alpha^2\sigma_L^2 + \beta^2\sigma_V^2}}{\alpha\sigma_L + \beta\sigma_V} \tag{54}$$

Let $r = (\alpha\sigma_L)/(\beta\sigma_V)$ then:

$$\delta = 2.73 \frac{\sqrt{1+r^2}}{1+r} \tag{55}$$

If, for example, $r = 1$, then $\delta \approx 1.93$, so that the circuit would need to be simulated (and its timing checked) with all its transistors' channel lengths and threshold voltage set at their $+1.93\sigma$ points. Notice that, since α and β depend on transistor sizing, then (55) provides a way in which δ can be controlled by circuit optimization and/or process tuning.

We have compared our yield upper and lower bounds to the yield Cauchy bounds derived in [11]. These Cauchy bounds were recomputed with path variance ratios equal to the ones used in this work, in order to have a valid comparison. Table 2 gives the different virtual corners δ for a desired yield of 95%: The first two columns list the values of δ predicted using our upper and lower bounds. Hence, for a circuit with unknown or uncertain within-die *positive* correlation, we predict a virtual corner of 1.43 in the best case and 1.93 in the worst case. Note that verifying timing using the latter corner will guarantee the desired yield of 95% for arbitrary positive correlation of the within-die variations.

The remaining columns list the "predicted" worst-case δ's using Cauchy lower bounds in [11] where within-die correlations are captured using PCA with order p. Depending on the extent of correlation of the within-die variations, p can increase with looser correlations, or decrease with tighter correlations. As shown in Table 2, as p increases from 2 to 8, δ also increases from 1.71 to 2.24. And for $p = 30$ (close to independent within-die variations), $\delta = 3.2$, which is overly pessimistic if compared to $\delta = 1.93$ in our work. It becomes clear that our approach, which is independent of the order of PCA p, gives better (smaller) predictions of δ in case correlation is unknown or uncertain, or in case correlation is known, but the PCA order is estimated to be larger than 4 ($p > 4$).

This comparison is further shown in Fig. 5, where our lower and upper bounds are plotted along with Cauchy lower bounds [11]. For any desired yield, the curves can be used to project a timing margin τ on the X-axis. Also, using (53) a virtual corner δ can be predicted from τ.

6. CONCLUSION

A method for statistical timing analysis has been developed, based on a timing yield model. The model is "full-chip" in that it can be applied with ease to large chips, in the pre-layout phase. It requires minimal process information (only parameter variances), and produces a "device file" setting or a "virtual corner" δ, that if used for all devices, and if timing is verified using traditional static timing analysis, the desired yield is achieved. Another dual result is worth noting. For a desired yield, a certain margin τ is predicted. Hence, if traditional static timing analysis with nominal device setting is used, and if the margin τ is allowed for every path, then circuit would pass timing verification with the desired yield.

Table 2: Virtual corners δ for 95% yield

	Upper bound	Lower bound	p = 2	p = 4	p = 6	p = 8
δ	1.43	1.93	1.71	1.93	2.1	2.24

Both the Setup (Max) and Hold (Min) yields are tackled in this paper. Upper and lower bounds are derived on these yields; these bounds hold for any *positive* correlation model of the systematic within-die variations. Correlations between gates on a path and between paths on a chip are studied in order to get the desired bounds.

7. REFERENCES

[1] M. Eisele, J. Berthold, D. Schmitt-Landsiedel, and R. Mahnkopf. The impact of intra-die device parameter variations on path delays and on the design for yield of low voltage digital circuits. *IEEE Trans. on VLSI*, 5(4):360–368, December 1997.

[2] A. Gattiker, S. Nassif, R. Dinakar, and C. Long. Timing yield estimation from static timing analysis. In *IEEE Int'l Symp. on Quality Electronic Design*, pages 437–442, San Jose, CA, March 26-28 2001.

[3] K. Singhal and V. Visvanathan. Statistical device models for worst case files and electrical test data. *IEEE Trans. on Semiconductor Manufacturing*, 12(4):470–484, November 1999.

[4] C. Visweswariah. Death, taxes and failing chips. In *Design Automation Conf.*, pages 343–347, Anaheim, CA, June 2-6 2003.

[5] J. A. G. Jess, K. Kalafala, W. R. Naidu, R. H. J. M. Otten, and C. Visweswariah. Statistical timing for parametric yield prediction of digital integrated circuits. In *Design Automation Conf.*, pages 932–937, Anaheim, CA, June 2-6 2003.

[6] A. Agarwal, D. Blaauw, V. Zolotov, and S. Vrudhula. Computation and refinement of statistical bounds on circuit delay. In *Design Automation Conf.*, pages 348–353, Anaheim, CA, June 2-6 2003.

[7] A. Devgan and C. Kashyap. Block-based static timing analysis with uncertainty. In *IEEE/ACM In'l Conf. on Computer-Aided Design*, pages 607–614, San Jose, CA, November 9-13 2003.

[8] S. Bhardwaj, S. B.K. Vrudhula, and D. Blaauw. tau: Timing analysis under uncertainty. In *IEEE/ACM Int'l Conf. on Computer-Aided Design*, pages 615–620, San Jose, CA, November 9-13 2003.

[9] H. Chang and S. S. Sapatnekar. Statistical timing analysis considering spatial correlations using a single PERT-like traversal. In *IEEE/ACM Int'l Conf. on Computer-Aided Design*, pages 621–625, San Jose, CA, November 9-13 2003.

[10] A. Agarwal, D. Blaauw, and V. Zolotov. Statistical timing analysis for intra-die process variations with spatial correlations. In *IEEE/ACM Int'l Conference on Computer-Aided Design*, pages 900–907, San Jose, CA, November 9-13 2003.

[11] F. Najm and N. Menezes. Statistical timing analysis based on a timing yield model. In *Design Automation Conference*, pages 460–465, San Diego, CA, June 7-11 2004.

[12] K. A. Bowman and J. D. Meindl. Impact of within-die parameter fluctuations on future maximum clock frequency distributions. In *IEEE Custom Integrated Circuits Conf.*, pages 229–232, 2001.

[13] S. G. Duvall. Statistical circuit modeling and optimization. In *Int'l Workshop on Statistical Metrology*, pages 56–63, June 2000.

[14] Y. L. Tong. *The multivariate normal distribution*. Springer-Verlag, New York, NY, 1990.

[15] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, NY, 2nd edition, 1984.