Analysis and Modeling of CD Variation for Statistical Static Timing

Brian Cline[†], Kaviraj Chopra[†], David Blaauw[†], Yu Cao[‡]

[†]University of Michigan, Ann Arbor, MI, [‡]Arizona State University, Tempe, AZ

{btcline,kaviraj,blaauw@umich.edu}　　　　{Yu.Cao@asu.edu}

Abstract

Statistical static timing analysis (SSTA) has become a key method for analyzing the effect of process variation in aggressively scaled CMOS technologies. Much research has focused on the modeling of spatial correlation in SSTA. However, the vast majority of these works used artificially generated process data to test the proposed models. Hence, it is difficult to determine the actual effectiveness of these methods, the conditions under which they are necessary, and whether they lead to a significant increase in accuracy that warrants their increased runtime and complexity. In this paper, we study 5 different correlation models and their associated SSTA methods using 35420 critical dimension (CD) measurements that were extracted from 23 reticles on 5 wafers in a 130nm CMOS process. Based on the measured CD data, we analyze the correlation as a function of distance and generate 5 distinct correlation models, ranging from simple models which incorporate one or two variation components to more complex models that utilize principle component analysis and Quad-trees. We then study the accuracy of the different models and compare their SSTA results with the result of running STA directly on the extracted data. We also examine the trade-off between model accuracy and run time, as well as the impact of die size on model accuracy. We show that, especially for small dies (< 6.6mm x 5.7mm), the simple models provide comparable accuracy to that of the more complex ones, while incurring significantly less runtime and implementation difficulty. The results of this study demonstrate that correlation models for SSTA must be carefully tested on actual process data and must be used judiciously.

1. Introduction

Static timing analysis (STA) has become a key method in the performance verification of modern chip designs and is the primary technique that abstractly incorporates manufacturing variation into design. Recently, the shortcomings of STA have become apparent with its inability to efficiently include within-die (intra-die) variation in process parameters such as gate length, oxide thickness, and doping levels. STA, in its most common form, is a case-based analysis: designers perform simulations given best-, nominal-, and worst-case conditions and all devices are assigned the same process parameter value. However, with continued process scaling past 65nm, within-die variation has become more prominent and also exhibits considerable spatial correlation. Intra-die variation tends to average out over the length of a circuit path and hence can reduce the circuit delay distribution. On the other hand, the presence of significant intra-die delay variation in two converging paths increases their maximum delay distribution. With a case based STA analysis, it is therefore difficult to construct a guaranteed bound on the actual timing distribution of a circuit without being overly conservative.

To address this issue, statistical static timing analysis (SSTA) was developed and received considerable attention in the CAD research community in recent years [1][3][5][6]. SSTA models process parameters, such as gate length and doping concentration, as random variables and propagates these random variables through the circuit in topological fashion, analogous to the propagation in its deterministic counterpart.

The first efforts in STA [2] modeled all process parameter variations, as well as the propagated arrival times, as independent random variables. This assumption significantly simplified the analysis but compromised accuracy. In [1][3][4], process parameters variations were still considered as independent, but correlations between arrival times due to reconvergence in the circuit were accounted for. In the latest generation of SSTA tools [1][5][6], correlations between the process parameters of different gates in the circuit were also accounted for.

Process parameter correlation exhibits both a die-to-die component (causing all device parameters in a die to vary by some common amount) and a spatially defined component (where devices with close proximity are more likely to have similar process parameter values). While die-to-die correlations can be incorporated relatively easily by enumerating a small number of die conditions, the so-called spatial correlations increase the complexity of SSTA substantially. Accounting for these correlations requires both a model which expresses the correlations in an amendable form, as well as an accompanying SSTA engine that can operate on this model.

The spatial correlation model proposed in [5] uses a grid-based approach where the process parameters of all gates that fall within the same grid square are assumed to be identical. The correlation between different grid squares is decomposed using principal component analysis (PCA), and then modeled as a weighted sum of independent random variables (the principle components). A different grid-based model was developed in [1]. Here, the authors combine multiple grids with varying granularity in a tree-like fashion, where each grid square is assigned an independent random variable and each gate is associated with every grid square in which it resides. While the Quad-tree uses a larger total number of random variables than the PCA approach (given the same grid granularity), less information is associated with each individual gate. One important item we address in this paper that was not included in [1] is a method for fitting the Quad-tree model to measured data.

Extensive work has been performed on modeling spatial correlations and improving the accuracy of SSTA using the discussed models. To our knowledge, however, no study has been published demonstrating the effectiveness of these methods on real process data. In fact, all above referenced SSTA methods used artificially constructed process data for verifying the accuracy of their approach. This is due, in part, to the lack of published data on process variation characteristics, such as the behavior of correlation as a function of spatial separation. However, without verification of the proposed models on actual process data, it is difficult to determine the accuracy of the approaches. Furthermore, it is possible that while the proposed methods are accurate, the correlation in actual process data is such

that even simpler models provide sufficient accuracy, making the complexity and runtime of the more complex methods unwarranted.

In this paper, we obtained measured critical dimension (CD) data through electrical linewidth measurement (ELM) of a 130nm test chip design. The test chip consists of 8 different test structures (various densities and orientations of polysilicon lines) [7] repeated at 308 sites per field over 23 fields and 5 wafers for a total of 35420 measurements. Note that this paper does not address CD variation due to layout. Since layout variation is deterministic, it can be included orthogonally to this work. Instead, the focus is on statistical variation and spatial correlation model testing. Specifically, the ELM data is used to study the correlation characteristics of actual CD variation as well as the effectiveness of different correlation models and their SSTA analyses.

Based on the measured data, different die sizes were extracted ranging from 6.6mm to 30.7mm on a side. For the smaller die sizes, multiple die could be positioned in a single field. Then, we analyzed general characteristics of the CD variation as well as the spatial correlation as a function of separation distance. We found that the correlation function differed significantly for vertical and horizontal separations, which had significant implications for correlation models. We then tested 5 different correlation models using the measured CD data and ran SSTA on these models. The 5 models include the PCA and Quad-tree methods, as well as three simpler methods that model the process parameters using a die-to-die and/or an independent variation component. For Quad-tree, we introduce a new fitting method which derives the model parameters from measured data.

The ability of the correlation models themselves to accurately represent the correlation structure was first studied by generating a large set of sample dies using Monte Carlo simulation on the correlation models and running STA on each generated sample die. In addition, SSTA was also performed directly on the generated models. By comparing these two circuit delay distributions with that obtained by running STA on the actual set of die (using the measured data), the accuracy of the different correlation models and the SSTA accuracy was studied. We also investigated the dependence of the analysis accuracy on the granularity of the models, their parameters, and the size of the die.

In our analysis, we found that both the PCA and Quad-tree approaches, combined with a Clark-based SSTA approach [8], resulted in an accurate estimation of the actual mean and variance of the circuit delay distributions. However, we also found that, especially for smaller die, their accuracy was not significantly higher than the accuracy obtained using the much simpler model which expresses the correlation using the sum of an inter-die component and random components. Hence, this study indicates that these more complex models may only be necessary for very large die and shows the importance of testing correlation models on actual process data and their judicious use.

The remainder of this paper is organized as follows: Section 2 describes the background information of different process variation sources and correlation models. Section 3 presents an analysis of the measured CD data and its correlations. In Section 4, the five correlation models are introduced and their effectiveness in SSTA is compared. The paper is concluded in Section 5.

2. Background

2.1 Process Parameter Variation

In today's fabrication processes, there are three process parameters that exhibit the most pronounced variation impact: doping, interlayer dielectric (ILD) thickness, and gate length.

2.1.1 Doping

The first major source of variation in today's circuits is caused by changing doping levels. The key mechanism behind this variation is known as random dopant fluctuation (RDF). However, unlike the other two process parameters, this variation is random and independent of other sources of variation [9]. Therefore, RDF can be modeled more easily than the other process parameters and does not require a spatial correlation model. Hence, we did not include it in our current analysis. However, the proposed analysis methodology can easily incorporate this additional source of variation.

2.1.2 Interlayer Dielectric Thickness

ILD variation affects interconnect parasitics and is caused by chemical-mechanical polishing (CMP) and variation in layout topography. However, since this is a relatively systematic variation that depends heavily on the density of the metallization in the design, it is possible to reduce this variation significantly or to model it deterministically.

2.1.3 Gate Length Variation

Out of the three main process parameters discussed here, gate length variation remains the dominant source of delay variation. This is due to the fact that there are numerous sources of gate length variation that affect the device at different stages throughout the fabrication process, such as etch rate variation, lithography dose variation, focus variation, lens imperfections, scanner variation, etc. [7]. Most of the gate length variation is due to imperfections in the lithographic system. These imperfections can contribute multiple levels of variation at any point during the fabrication process. For example, changes in the wafer machinery over long periods of time can introduce wafer-to-wafer variation. Also, within-wafer variation originates from non-uniformities in temperature, laser intensity fluctuations or film-thickness changes [7]. Below the scale of intra-wafer variation lies reticle variation, which can be caused by deviations in dose and focus. Lastly, the optical system can induce within-die variation if there are imperfections in the lens as the scanner moves through a reticle.

Most of the sources of gate length variation are systematic in nature, and previous work has shown that it is possible to remove these systematic variations, given sufficient information about the topology surrounding a device [7]. Hence, in this work, we focus on the random component of variation (as opposed to its systematic component), which can exhibit spatial correlation.

2.2 Types of Gate Length Variation

Within the random component of gate length variation, we can further distinguish three types of variation: independent, die-to-die, and spatially correlated. For this section all variables ΔL_x are assumed to be zero mean, unit variance random variables.

2.2.1 Independent

In this case, each device in the design has variations in its process parameters that are independent from the variations in process parameters of other devices. Independent variations can be modeled using independent random variables. If the gate lengths in a die are completely specified by independent variations, the length of gate i can be expressed as follows:

$$L_{g,i} = L_{nom,i} + \sigma_{ri}\Delta L_{rnd,i}, \tag{1}$$

where $L_{nom,i}$ is the nominal value of gate length for that gate, $\Delta L_{rnd,i}$

is the random device length variation for gate i, and σ_{ri} is the sensitivity of gate i to changes in $\Delta L_{rnd,i}$.

2.2.2 Die-to-Die

Die-to-die variation, on the other hand, describes variation that is common for all devices on a particular die. When only inter-die variation is considered, all gate lengths within a particular die become perfectly correlated. Therefore, the gate length of gate i, only considering die-to-die variation, can be expressed as:

$$L_{g,i} = L_{nom,i} + \sigma_{dd}\Delta L_{die-to-die}, \tag{2}$$

where $L_{nom,i}$ is the nominal value for gate i, $\Delta L_{die-to-die}$ is a single random variable that is applied to all gates in the circuit, and σ_{dd} is the global gate sensitivity to changes in $\Delta L_{die-to-die}$.

2.2.3 Spatially Correlated

The last type of variation that we consider is spatially correlated variation. Most process variation within a single die is spatially correlated, and generally, correlation decays as a function of distance between two points. Generally, in statistical timing analysis, the desire is to express correlation using a weighted sum of independent random variables, as shown below,

$$L_{g,i} = L_{nom,i} + \alpha_1\Delta L_1 + \alpha_2\Delta L_2 + \alpha_3\Delta L_3 + \dots, \tag{3}$$

where ΔL_k is the variation of the kth component and α_k is the sensitivity of the gate length to changes in the kth component. By maintaining this form throughout the timing analysis, correlation information between the arrival times can be maintained. The specific values of the sensitivities and the number of components will vary between the different correlation models, which are discussed in the following subsection.

2.3 Correlation Models

As briefly mentioned in the introduction, the five correlation models analyzed in our experiments are: die-to-die (D2D), independent (commonly referred to as random and hence forth referred to as such in this paper), D2D + random, PCA, and Quad-tree.

2.3.1 D2D, Random, and D2D + Random

The equations used to express the length variation of a particular gate for the die-to-die and random cases are equal to (1) and (2) discussed in the previous section. Therefore, the die-to-die + random variation is a combination of (1) and (2):

$$L_{g,i} = L_{nom,i} + \sigma_{dd}\Delta L_{die-to-die} + \sigma_{ri}\Delta L_{rnd,i}, \tag{4}$$

Once we know the forms of these equations, it is simple to develop sensitivity matrices, which are the input to our statistical timing tool. For instance, the sensitivity matrices for D2D, random, and D2D + random are shown in (5) as D, R, and DR, respectively.

$$R = \begin{bmatrix} \sigma_{r1} & 0 & 0 & \dots & 0 \\ 0 & \sigma_{r2} & 0 & \dots & 0 \\ 0 & 0 & \sigma_{r3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_{rm} \end{bmatrix} \quad D = \begin{bmatrix} \sigma_{dd} & \sigma_{dd} & \sigma_{dd} & \dots & \sigma_{dd} \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

$$DR = \begin{bmatrix} \sigma_{dd} & \sigma_{dd} & \sigma_{dd} & \dots & \sigma_{dd} \\ \sigma'_{r1} & 0 & 0 & \dots & 0 \\ 0 & \sigma'_{r2} & 0 & \dots & 0 \\ 0 & 0 & \sigma'_{r3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma'_{rm} \end{bmatrix}, \tag{5}$$
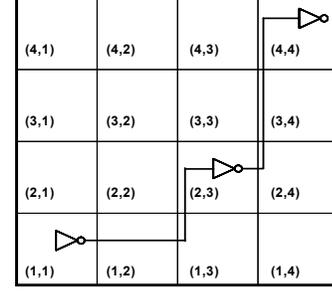

Figure 1. PCA Grid Example

where σ_{dd} is the standard deviation of only the die-to-die component, σ_{ri} is the standard deviation of the ith random component, and σ'_{ri} is the standard deviation of the ith random component when the die-to-die component has been removed.

2.3.2 PCA

The PCA model is a grid-based model (shown above in Figure 1) that separates the die into n grids. Each grid is associated with a principal component, and all n principal components are independent, normal random variables with zero mean and unit variance. Here, PCA deals with spatially correlated distributions, so we know that its equation is going to look similar to (3). In fact, for a gate i, its length can be expressed as:

$$L_{g,i} = L_{nom,i} + \sum_j \alpha_{ij}\Delta L_j, \text{ where } \alpha_{ij} = \sigma_i v_{ij}\sqrt{\lambda_j}, \tag{6}$$

where ΔL_j is the jth component and α_{ij} is calculated as stated in (6) – σ_i is the standard deviation associated with grid i, v_{ij} is the ith element in the jth eigenvector of the correlation matrix, and λ_j is the jth eigenvalue of the correlation matrix [5]. Therefore, the sensitivity matrix, P, for the PCA model will be of the form,

$$P = \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} & \dots & \alpha_{1,m} \\ \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} & \dots & \alpha_{2,m} \\ \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} & \dots & \alpha_{3,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{n,1} & \alpha_{n,2} & \alpha_{n,3} & \dots & \alpha_{n,m} \end{bmatrix}, \tag{7}$$

where each grid is associated with one column and one row.

2.3.3 Quad-Tree

Quad-tree is another grid model that utilizes various grid levels combined in a tree-like structure – shown in Figure 2 – to include spatial correlation. The Quad-tree has $l+1$ levels, and each level, k, contains 2^k-by-2^k squares [1]. Levels are numbered where "level 0" represents the top level and l is bottommost level. Level 0 only has one grid, while level k has 4^k grids. All of the regions at different levels of the tree are associated with an independent random variable that includes part of the total intra-die variation. For a gate located within bottommost region r, the associated variation is a sum of all the intra-die variation components that intersect region r as you progress up the tree (i.e. in Figure 2 grid (2,13) intersects grid (1,3) and (0,1)). For example, the equation for gate length for a gate that lies in grid (2,7) is,

$$L_{g,(2,7)} = L_{nom,(2,7)} + \Delta L_{(2,7)} + \Delta L_{(1,2)} + \Delta L_{(0,1)}. \tag{8}$$
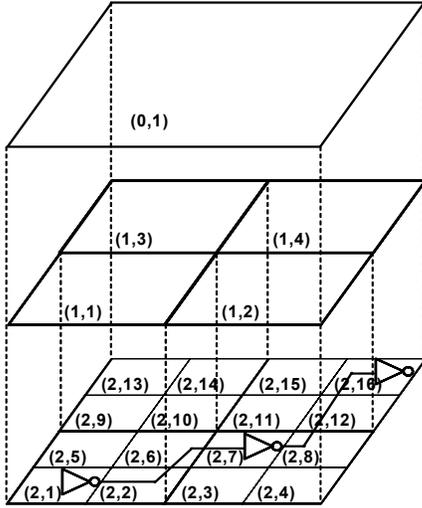
Figure 2. Quad-tree Model Example

Figure 3. Quad-tree Model Example

Thus, the sensitivity matrix is similar to the PCA's matrix in (7), where all grids (including all levels of the tree) are given one row in the sensitivity matrix, while only the bottommost grids are assigned to individual columns. All grids that do not intersect with a particular bottommost grid (assigned column i in the matrix) will have a zero sensitivity value at its row j in the matrix (i.e., element $[i, j]$ equals zero).

As stated in the introduction, the authors in [1] did not explain how to fit actual data to their model. After examining several different algorithms, we derived a Quad-tree fit that is efficient, simple and provides good accuracy. Prior to fitting, we discovered that closely matching the die-to-die component was very important to overall accuracy, therefore, this fitting method was designed to accurately capture the die-to-die component first with zero error.

Essentially, the fitting method starts at level 0 and works its way down the tree. At each level you stop there, determine how many grids you have at that level (4^k for grid level k). Next, you parse through each grid at that level and calculate its mean. You do this for all dies across all reticles and wafers and then calculate the standard deviation of grid mean for that particular grid. This is what goes in your sensitivity matrix. As stated previously, do this for all grids in a level, then move to the next level down.

The pseudo code for the fitting algorithm is as follows:

```
Begin at level = 0.
while (level < total number of levels)
    grid = 1.
    while (grid < 4 ^ level)
        compute grid μ
        compute σ_i of grid i mean for all dies
        enter σ_i into sensitivity matrix
    end
end
```

3. Experimental Data and Analysis

As stated earlier, our analysis is based on 130nm ELM data taken from horizontal polysilicon lines that have optical proximity corrections (OPC) included [7]. We investigated 5 different wafers, each wafer contained 23 fields, and each field included 308 measurement points: 14 points in the horizontal direction and 22 points in the vertical direction. Individual measurement points were spaced horizontally by 2.19mm and vertically by 1.14mm.
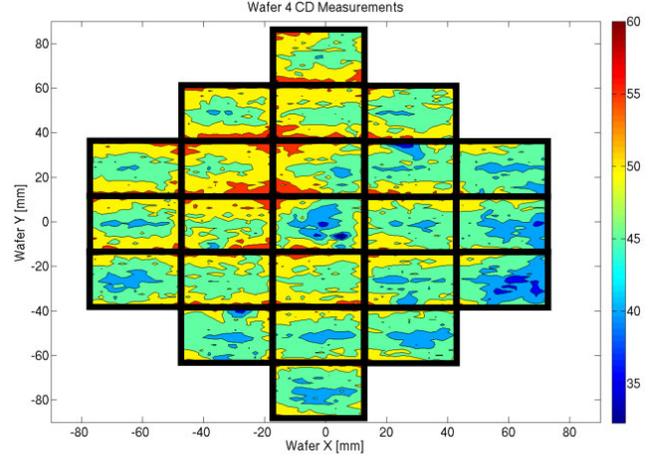
An example of one wafer of ELM CD measurements is shown in Figure 3. As shown, not only do the measurements vary across the wafer (the lower right corner has smaller CD values than the upper right corner), but specific patterns occur within the reticles (the upper and lower boundaries of the field have a higher CD than the center points.) For these 5 wafers, we divided the reticles into various die sizes in order to investigate the effects of die size on the CD variation. Initially, we diced a reticle into 4 die, (a 2-die by 2-die configuration where each die is 15mm x 12mm). Then, we examined a number of characteristics including the mean, standard deviation, and correlation of all the dies.

The mean values for each data point in a die using a 2 x 2 diced reticle are shown in Figure 4 (a). As can be seen, the typical die has lower values in the center of the die, and CD values increase as you approach the edges of the die. In Figure 4 (b), the standard deviation over mean is plotted for this particular case. Again, the figure shows
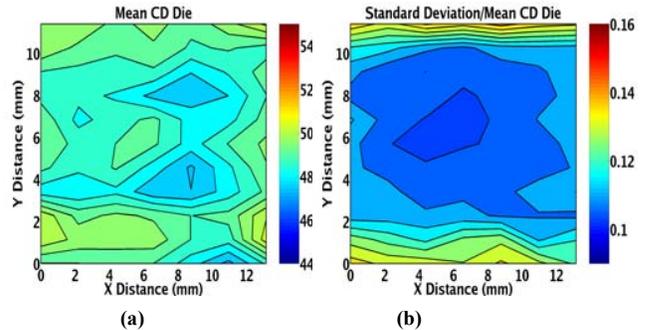

(a)   **(b)**

Figure 4. (a) Mean CD Values for Die (2x2 reticle dice)
(b) Standard Deviation/Mean for Die (2x2 reticle dice)
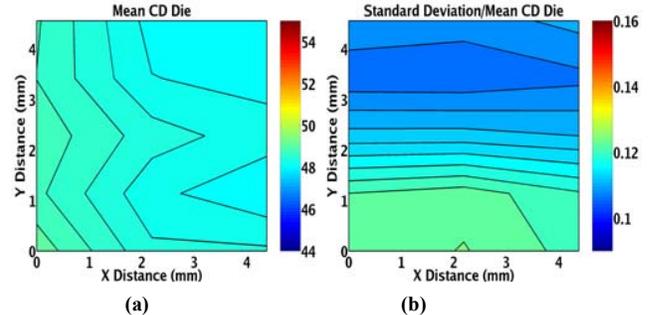

(a)   **(b)**

Figure 5. (a) Mean CD Values for Die (4x4 reticle dice)
(b) Standard Deviation/Mean for Die (4x4 reticle dice)

Figure 6. Average Correlation vs. Distance

Figure 8. PDF Plot for ELM Measured CD
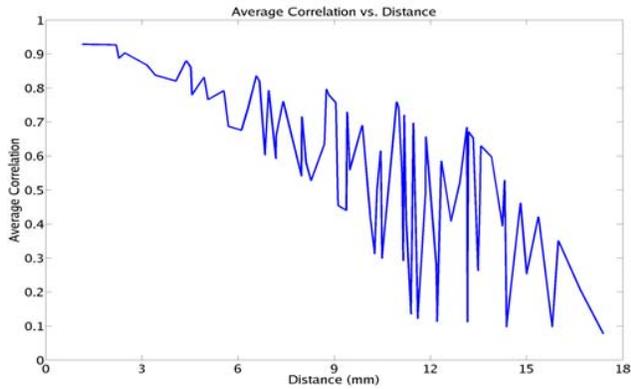
the edge effects in the die. To contrast the 2 x 2 diced reticle, we have also included the 4 x 4 diced versions in Figure 5.

On average, the 4 x 4 dicing merely divides the 2 x 2 case into two-by-two grids of its own. Thus, it can be seen that the 4 x 4 mean plot is a quarter of the 2 x 2 plot, with the spot effect seen in the 2 x 2 case lying on the inner portion of the 4 x 4 die. Similarly, the standard deviation-over-mean plot also resembles a quarter of the 2 x 2 case, with the lower deviation occurring at the top edge of the typical 4 x 4 die. It should be noted, however, that the variation structure is quite different between the 2 x 2 and 4 x 4 diced cases.

In addition to the mean and standard deviation, the correlation was also extracted for different size die. Plotted in Figure 6 is the average correlation versus separation distance. It is easily identifiable that this function is not monotonically decreasing with distance, x. On the contrary, we see many distinctive peaks where correlation falls and then spikes up at a particular distance. From this investigation, it became clear that correlation versus horizontal distance is different from the correlation versus vertical distance (i.e. correlation is typically stronger along a particular axis). This is confirmed in Figure 7 (a) where correlation is plotted separately for spacing in horizontal and vertical directions.

As shown, correlation in the x-direction is actually stronger which can be explained by the fact that during fabrication only a narrow slit in the x-direction is being printed while the entire y-dimension of the reticle is being printed. Thus, vertically, the reticle sees all of the variation in the lithographic system (particularly lens aberrations) but as the reticle is scanned across x, this variation does not change significantly, creating higher correlation in the x direction. Figure 7 (b) also shows similar behavior for a much smaller die size.
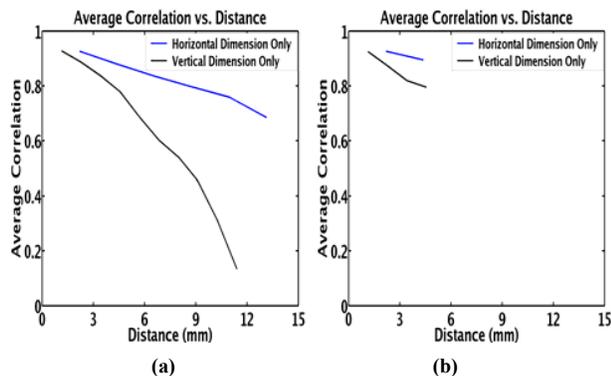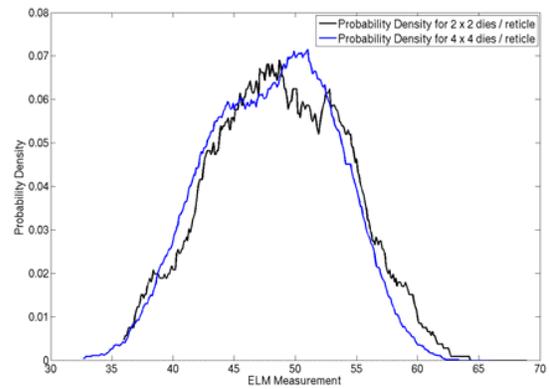
Lastly, we plotted probability density functions for each point within a die. One example is shown in Figure 8, which is a plot of point 76 within the 15mm x 12mm die (2x2 reticle) and point 14 within the 7mm x 6mm die (4x4 reticle).

4. Variation Modeling and SSTA Results

After analysis of the experimental data, we used the data to test the accuracy of different correlation models and their associated SSTA runs. For our test circuit, we utilized the behavioral Verilog from an industrial, 15000 gate implementation of a Viterbi decoder. Then Synopsys's Design Compiler was used to synthesize the design and balance the paths. Lastly, the test circuit was placed and routed using Cadence's Silicon Ensemble, in order to generate the placement information needed by the SSTA tool. The authors would like to note that we did not actually layout the Viterbi decoder. It was merely used as a simulation benchmark to test the accuracy of our spatial correlation models.

The general flow of our analyses is illustrated in Figure 9. There are effectively three branches in the flow. All branches start with the same wafer data. Then, in the first case (the left branch), we perform static timing analysis on all N die, where,
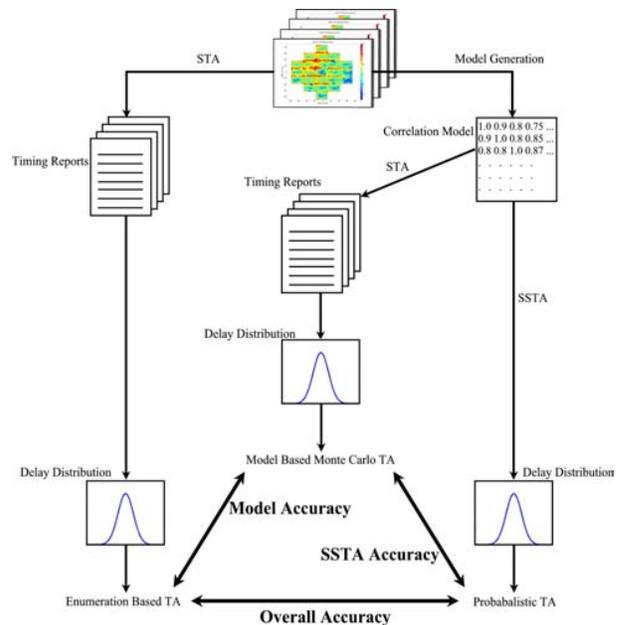

Figure 7. Average Correlation vs. Distance (1-dimension only)
(a) 2x2 reticle dice (b) 4x4 reticle dice

Figure 9. Timing Analyses Flow

Table 1. Enumeration-Based, Model-Based, and Probabilistic TA Results

Analysis Method		μ (ns)	% Error from Enum.	σ (ns)	% Error from Enum.
Enumeration-based TA		2.049		0.152	
Model-based Mont Carlo TA	Die-to-Die	1.934	5.623%	0.139	8.326%
	Random	2.087	1.849%	0.058	62.12%
	D2D + Random	2.006	2.117%	0.146	3.784%
	PCA	2.033	0.800%	0.151	0.428%
	2-level Quad-tree	2.006	2.111%	0.159	4.556%
Probabilistic TA	Die-to-Die	1.945	5.108%	0.146	3.789%
	Random	2.130	3.934%	0.040	73.70%
	D2D + Random	2.006	0.769%	0.146	3.793%
	PCA	2.071	1.043%	0.148	2.694%
	2-level Quad-tree	2.061	0.577%	0.157	3.198%

$$N = X \times Y \times 23 \times 5 , \qquad (9)$$

X is the number of die per reticle in the horizontal direction, and Y is the number of die per reticle in the vertical direction (23 is the number of reticles, 5 is the number of wafers). From deterministic STA, we obtain N timing reports from which we can extract a final distribution for critical path delay of the circuit. We consider this our golden statistical analysis of the circuit delay since it is based directly on the underlying measured data.

The two paths on the right both begin with a model generation step which involves fitting one of the 5 discussed models to the data. These two paths then diverge. The center path (referred to as model-based Monte Carlo) essentially follows a similar analysis as the enumeration-based timing analysis (TA), with the major difference that the STA is performed on random die generated based on the fitted model instead of the actual measured data. The final left path performs SSTA directly on the fitted correlation model.

In the end, this TA flow gives us three outputs available for comparison. By comparing the enumeration-based TA with the model-based Monte Carlo TA, we can determine the inherent accuracy of the correlation model for timing analysis. Similarly, by comparing the probabilistic TA distribution to the model-based Monte Carlo TA, the accuracy of SSTA run on a particular correlation model can be determined. Lastly, we can determine the overall accuracy of using SSTA with a specific correlation model by directly comparing the enumeration-based TA to the probabilistic TA.

Table 1 includes the results of our TA verification flow. Included are all three TA flow outputs: enumeration-based, model-based Monte Carlo, and probabilistic. Additionally, Figure 10 shows sample probability density plots for 3 of the models, including the enumeration-based and two PCA models (model-based and probabilistic). All of the curves in Figure 10 are from the 4 x 4 reticle dice experiment.

When examining the model-based Monte Carlo TA results in Table 1, it is clear that even the simple die-to-die models only deviate from the enumeration-based results by less than 10%. The random model is more accurate than die-to-die with regards to the mean, but is significantly off with regards to standard deviation. The reason for this is that die-to-die variation tends to produce variation in the circuit delay where as random or spatially correlated variations tend to average out over the circuit paths and therefore instead result in a shift of the mean value of circuit delay. Since the random correlation
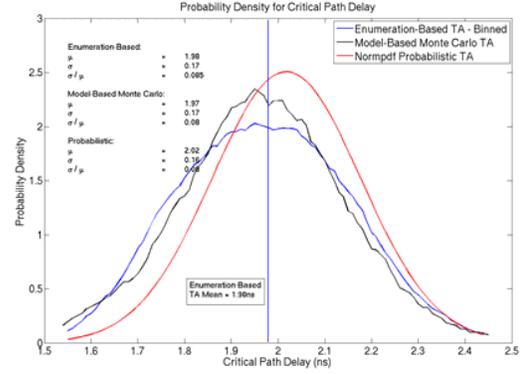

Figure 10. Probability Density Plots for 3 Models (Enumeration-Based, PCA Model-Based Monte Carlo, and PCA Probabilistic)

model does not model die-to-die variation at all, it incurs a significant error in the standard variation of the circuit delay. The "Die-to-die + random" correlation model improves over die-to-die due to the fact that it models what appear to be the two most important components of the variation. Overall, it is clear from the simple models that both random (independent) and die-to-die variation are significant to the accuracy of the model, but die-to-die is substantially more significant.

Table 1 also shows the two more complex spatial models for model-based Monte Carlo TA. The error in PCA is negligible, falling below 1%, while that of the Quad-tree is somewhat higher. The fact that the PCA correlation model out-performs the Quad-tree for model based Monte Carlo TA is not surprising since it utilized a much large number of principle components to fit the measured data.

One of the more surprising results is that when we look at the error of the probabilistic TA for the 5 models, PCA and Quad-tree report very comparable results despite the fact that PCA reported better results for model based Monte Carlo TA. Both are less than 1% away from the means of the enumeration-based run, and approximately 3% off from the standard deviations. Hence, the PCA model is more accurate than the Quad-tree model, but the execution of SSTA on PCA incurs more error than SSTA execution on the Quad-tree, making the final results approximately equal. This behavior was also seen consistently for different tests that were performed. We postulate that this behavior is the result of the large number of independent components associated with each gate in PCA. This flexibility allows PCA to obtain a better fit of the data in its correlation model, but also makes the SSTA task more difficult and introduces a higher error in the Clark based max function that is performed inside the SSTA tool. Finally, it is noteworthy that the simple die-to-die + random model performs nearly as well on the probabilistic TA flow as the more complex models.

4.1 Model Accuracy vs. Die Size

Next, we studied the affect that die size had on the models and SSTA accuracies. That result is shown above in Table 2. The cells in the first row contain the enumeration-based results for mean and standard deviation, while the rest of the table displays the percent deviation from the enumeration-based TA. In general die-to-die, die-to-die + random, and Quad-tree models become more accurate (in terms of overall accuracy) as the die gets smaller. This is intuitive in terms of the die-to-die model because by shrinking the die, we are making more of the variation inter-die variation, and since we fit the Quad-tree to die-to-die variation first, it also follows this trend.

Table 2. Model versus Die Size

Run Type		23mmx19mm (1.2x1.2 reticle dice)		15mmx13mm (2x2 reticle dice)		8mmx6mm (4x4 reticle dice)	
		μ (ns)	σ (ns)	μ (ns)	σ (ns)	μ (ns)	σ (ns)
Enumeration-based TA		2.022	0.156	2.049	0.152	1.975	0.167
Model-based Monte Carlo TA	Die-to-Die (D2D)	4.176%	6.733%	5.281%	2.138%	2.407%	2.405%
	Random	2.136%	68.176%	1.772%	62.396%	4.545%	51.130%
	D2D + Random	0.029%	3.605%	1.105%	3.050%	0.103%	2.799%
	PCA	0.271%	6.259%	0.303%	3.472%	0.315%	1.209%
	1-level Quad-tree	3.165%	6.131%	3.098%	0.239%	0.173%	4.542%
	2-level Quad-tree	0.873%	8.979%	1.056%	1.688%	0.675%	2.039%
Probabilistic TA	Die-to-Die (D2D)	3.825%	8.492%	5.108%	3.789%	1.469%	3.192%
	Random	3.176%	83.625%	3.934%	73.703%	8.841%	62.472%
	D2D + Random	1.245%	11.247%	0.767%	3.793%	1.585%	3.188%
	PCA	0.099%	8.049%	1.043%	2.694%	2.138%	4.740%
	1-level Quad-tree	2.468%	7.451%	1.549%	1.424%	0.341%	0.280%
	2-level Quad-tree	0.794%	7.326%	0.027%	1.983%	1.002%	0.069%

The random model, on the other hand, actually becomes less accurate as die size decreases because it models all within-die variation as uncorrelated, and the dies will, on the whole, become more correlated as you shrink them due to the inverse relation between correlation and distance.

The last model, PCA, has a minimum in terms of accuracy. For large die (i.e., die that are larger than one-quarter of the reticle) and small die (like the 4x4 reticle case) PCA does not do as well, where as for medium size die it is more accurate. The results show that the relative model accuracy changes based on die size and hence, different models may be appropriate for different die size scenarios.

4.2 Grid Model Behavior

The way in which PCA and Quad-tree behaved while varying their characteristics, such as the number of principal components for PCA and number of tree levels for the Quad-tree, was also investigated.

First of all, limiting the number of principal components within PCA is commonly used since principal components are arranged in order of decreasing importance. For our purposes, we investigated the minimum number of principal components needed to obtain accurate results from SSTA. The behavior of the mean and standard deviation of SSTA versus number of principal components is given in Figure 11, and both are normalized to their respective value that includes all principal components. As you can see, both curves flat-ten out around 3 principal components, and approach 1 as the number of principal components gets large

Also of interest was the number of levels included in Quad-tree. However, for the tests that we ran, any number of levels above 3 did not give much additional gain in accuracy, since the Quad-tree SSTA already had errors of <1% for means and ~1% errors in standard deviation, as compared to the enumeration based model.

5. Conclusion

In our analyses, we found that the grid-based models were superior, both in the model-based simulations as well as the probabilistic TA. On average, Quad-tree was consistently more accurate with respect to the mean, and it outperformed PCA when the die size was small (6mm x 5mm). However, in all cases, the "die-to-die + random" model was only a couple percent behind Quad-tree and PCA in terms of error. Thus, the "die-to-die + random" model is a much simpler method of getting in the same ballpark as PCA and Quad-tree if you can tolerate a certain amount of error.

6. Acknowledgements

7. References

[1] A. Agarwal, D. Blaauw, V. Zolotov, S. Sundareswaran, M. Zhao, K. Gala, R. Panda, "Statistical delay computation considering spatial correlations," ASP-DAC 2003.
[2] M. Berkelaar, "Statistical Delay Calculation, a Linear Time Method," Proceedings of TAU97, Austin, TX, December 1997.
[3] J.J. Liou, et. al, "Fast Statistical Timing Analysis By Probabilistic Even Propagation," DAC 2001.
[4] A. Devgan, and C. Kashyap, "Block-based static timing analysis with uncertainty," IEEE/ACM ICCAD, November 2003.
[5] H. Chang and S.S. Sapatnekar, "Statistical timing analysis under spatial correlations," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Volume 24, Issue 9 2005.
[6] H. Chang, V. Zolotov, S. Narayan, C. Visweswariah, "Parameterized block-based statistical timing analysis with non-Gaussian parameters, nonlinear delay functions," Proceedings of the 42nd Design Automation Conference, June 2005.
[7] J. Cain and C.J. Spanos, "Electrical linewidth metrology for systematic CD variation characterization and causal analysis," Proceedings of SPIE Int. Soc. Opt. Eng. 2003.
[8] S.D. Samaan, "The impact of device parameter variation on the frequency and performance of VLSI chips," ACM/IEEE ICCAD 2004.
[9] C.E. Clark, "The greatest of a finite set of random variables," Operations Research, Volume 9, pp.85-91, 1961.
[10] P. Friedberg, et. al, "Modeling Within-Field Gate Line Length Spatial Variation for Process-Design Co-Optimization," ISQED, March 2005.
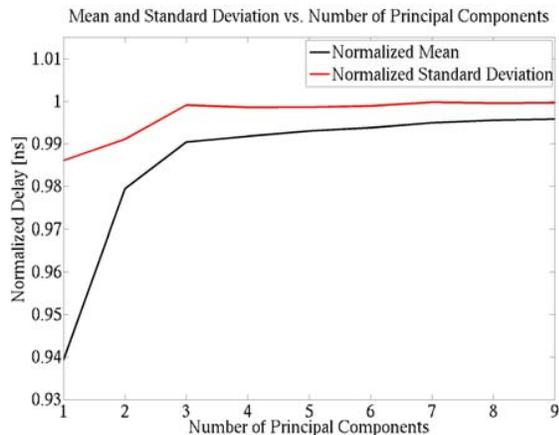

Figure 11. Mean and Standard Deviation versus Number of Principal Components (Normalized to Probabilistic Mean and Standard Deviation with all Principal Components)