Physically Justifiable Die-Level Modeling of Spatial Variation in View of Systematic Across Wafer Variability

Lerong Cheng, Puneet Gupta, *Member, IEEE*, Costas J. Spanos, *Fellow, IEEE*, Kun Qian, *Student Member, IEEE*, and Lei He, *Senior Member, IEEE*

***Abstract*— Modeling spatial variation is important for statistical analysis. Most existing works model spatial variation as spatially correlated random variables. We discuss process origins of spatial variability, all of which indicate that spatial variation comes from deterministic across-wafer variation, and purely random spatial variation is not significant. We analytically study the impact of across-wafer variation and show how it gives an appearance of correlation. We have developed a new die-level variation model considering deterministic across-wafer variation and derived the range of conditions under which ignoring spatial variation altogether may be acceptable. Experimental results show that for statistical timing and leakage analysis, our model is within 2% and 5% error from exact simulation result, respectively, while the error of the existing distance-based spatial variation model is up to 6.5% and 17%, respectively. Moreover, our new model is also 6X faster than the spatial variation model for statistical timing analysis and 7X faster for statistical leakage analysis.**

***Index Terms*— SSTA, spatial correlation, timing analysis, leakage analysis, yield modeling**

I. Introduction

With the CMOS technology scaling, process variation has become a major concern for VLSI design. Modeling and analyzing process variation has attracted a lot of attention.

Several works focus on analyzing and modeling of process variation [1]–[8]. The simplest method models process variation as the sum of inter-die (global) variation and independent within-die (local random) variation [4]. Later, it was observed that within-die variation is spatially correlated and the correlation depends on the distance between two within-die locations. [1], [8] model spatial variation as correlated random variables, and principle component analysis is applied to perform statistical timing analysis. In this model, a chip is divided into several grids and each grid has its own spatial variation. The spatial variations of different grids are correlated and the correlation coefficient depends on the distance between two grids. [2] focuses on the extraction of spatial correlation and it models the correlation coefficient as a function of distance. Several more complex spatial correlation models have been proposed in [9]–[13].

In contrast to the spatial correlation models, process oriented modeling has concluded that within-die spatial variation is caused by deterministic across wafer and across-field variation while purely random within-die spatial variation is not significant [14]–[16]. However, in practical design flow, designers do not know the within-wafer location or within-field location of each die; therefore, we need to analyze the impact of across-wafer variation and across-field variation on die-scale. Since silicon measurements cited in this paper indicate that across-wafer variation is much more significant than the across-field variation, we consider only across-wafer variation in this paper, but the approach is easily extended to account for across-field variations.

In this paper, we first analyze the impact of deterministic across-wafer variation on spatial correlation. We observe that when quadratic across-wafer variation model is used as in [15], [17], [18]:

1) Different locations on the chip may have different mean and variance. Such differences increase when the chip size increases.
2) When chip size is small, the correlation coefficients for a certain Euclidean distance are within a narrow range. This explains why most existing works find that spatial correlation is a function of distance.
3) Within-die spatial variation is *NOT* spatially correlated when across-wafer systematic variation is removed.
4) Within-die spatial variation is *NOT* independent from inter-die variation.
5) If chip size is small enough, the two-level inter-/within-die decomposition of process variation is still very accurate.

Based on our analysis, we propose three accurate and efficient spatial variation models[1] considering across-wafer variation. Experimental results show that our model is more accurate and efficient compared to the distance-based spatial

Manuscript received February 4, 2010; revised June 7, 2010 and September 23, 2010. This work was supported in part by Integrated Modeling Process and Computation for Technology (IMPACT). This paper was recommended by Associate Editor Luis Miguel Silveira.

L. Cheng , P. Gupta, and L. He are with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA (email {lerong, puneet, lhe}@ee.ucla.edu).

C. J. Spanos and K. Qian are with the Department of Electrical Engineering and Computer Science, Unveristy of California, Berkeley, CA 94720-1770 USA (email {spanos, qiankun}@eecs.berkeley.edu).

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

Digital Object Identifier.

[1]The programm and data of our proposed model can be downloaded in http://nanocad.ee.ucla.edu/Main/Stat

variation model in [2]. Compared to the exact simulation, error of our model for statistical timing analysis is within 2% and the error for statistical leakage analysis is within 5%. On the other hand, the error of the distance-based spatial correlation model is up to 6.5% for statistical timing analysis and up to 17% for statistical leakage analysis. Moreover, our model is 6X faster than the distance-based spatial correlation model for statistical timing analysis and 7X faster for statistical leakage analysis.

The rest of this paper is organized as follows: Section II discusses the physical causes for across-wafer variation; Section III analyzes the impact of across-wafer variation on die-scale; Section IV discusses the case when the across-wafer variation is not a perfect parabola; Section V introduces the new variation models; the new models are applied to statistical timing analysis in Section VI and statistical leakage analysis in Section VII; Section VIII summarizes the advantages and disadvantages of different variation models; Section IX further discusses the case when the across-wafer variation is an arbitrary function; and finally Section X concludes this paper.

II. Physical Origins of Spatial Variation

In silicon manufacturing, there are many steps that cause non-uniformity in devices across the wafer. Interestingly, most of these processes by the very nature of the equipment follow a radially varying trend across the wafer. Most processes are "center-fed" or "edge-fed" with the boundary conditions at the edge of wafer being substantially different. Moreover, wafers are often rotated to increase process uniformity across them which further leads to radial behavior of non-uniformity. This is further exacerbated by advent of single-wafer processing for 300mm wafers.

For example, overlay error includes errors in the position and rotation of the wafer stage during exposure, wafer stage vibration, and the distortion of the wafer with respect to the exposure pattern [19]. Magnification and rotation components of overlay error increase from center of the wafer outwards.[2] During chemical vapor deposition (CVD) step, species depletion and temperature non-uniformity on the wafer at lower temperatures may cause thickness non-uniformity [20], [21]. Redeposition effect in physical vapor deposition (PVD) [22] may cause non-uniformity of etch rate. Moreover, center peak shape of the RF electric field distribution [23] also leads to a center peak shape of etch rate, and chamber wall conditions [24] also cause etch rate non-uniformity. In real processes, the wafers are rotated to improve uniformity. [22], [24] show that the etch rate varies radially across the wafer: the etch rate is high at the center of the wafer and decreases toward the edges. Post-exposure bake (PEB) temperatures are higher at the center of the wafer and decreases outwards [25]. Similarly, other processes ranging from resist coat to wafer deformation due to vacuum chuck holding it follow a bowl-shaped trend across the wafer. All these processes cause a systematic across-wafer variation in physical dimensions.

Across-wafer variation of gate length observed in several recent silicon measurements [15], [17], [18], [26] validates our arguments. [27] also shows that ring oscillator frequency and leakage current decrease from the center to the edge of the wafer. Figure 1 shows industrial data of ring oscillator frequency for wafers from two different industrial processes. Process 1 is with 45nm technology and process 2 is with 65nm technology. From the figure, we see that for both process, ring oscillator frequency decreases from the center to the edge of the wafer. Moreover, it has also been shown that there is no spatial correlation for threshold voltage variation [14]. Therefore, the across wafer frequency and leakage variation is mainly caused by gate length variation.

[2]Overlay error can directly impact critical dimension in double patterning.

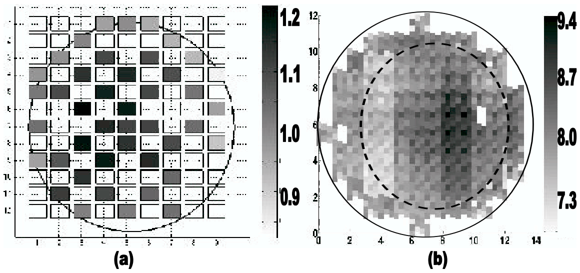

Fig. 1: Ring oscillator frequency within a wafer. (a) Process 1; (b) Process 2.

It has been shown that for process 1, the across-wafer frequency variation can be approximated as a quadratic function (a parabola) [27]. For process 2, the across-wafer variation is not a perfect parabola as process 1. However, it follows a systematic trend that the ring oscillator frequency decreases from the center to the edge of the wafer. Since the measurement data for process 2 (more than 300 wafers) is much more than process 1, in the rest of this paper, all of our simulation and experiments are based on the measurement result of process 2.

Besides across-wafer variation, lithography-induced effects such as lens aberrations can lead to systematic across-field variation and across-die variation. Across-die variation can be modeled as within-die deterministic mean shift and will not cause within-die spatial correlation. Moreover, silicon measurements cited in this paper indicate that across-wafer variation is much more significant (probably due to advancements in resolution enhancement and lithographic equipment) than across-field and across-die variation. Hence, for simplicity, we consider only across-wafer variation in this paper.

III. Analysis of Wafer Level Variation and Spatial Correlation

In this paper, a variation source V, such as L_{eff}, is be modeled as:

$$V = v_0 + v_c + v_p \tag{1}$$

where v_0 is the nominal value, v_c is a systematic constant offset, and v_p is the uncertainty part of process variation. Since both v_0 and v_c are constant, we may combine them as one constant term. The uncertainty term v_p is modeled as:

$$v_p = v_{aw} + v_{d-d} + v_{ad} + v_{af} + v_r \tag{2}$$

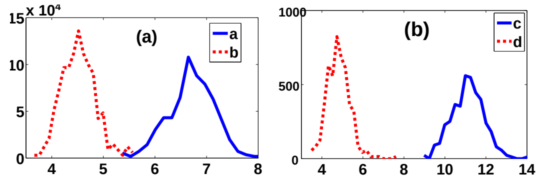

Fig. 2: PDF of across-wafer variation coefficients. (a) PDf of a and b. (b) PDF of c and d.

v_{d-d} comprises of inter-die random, inter-wafer, inter-lot variation and fitting error [3] of quadratic fitting of across wafer variation; v_{af} and v_{ad} are the across-field and across-die variation, respectively. As discussed in Section II, we consider only across-wafer variation and ignore these two types of variations (v_{af} and v_{ad}) in this paper; v_r is the random noise; v_{aw} is across-wafer variation, which is modeled as a quadratic function as in [3], [15], [17], [18]:

$$v_{aw}(x_w, y_w) \quad = \quad ax_w^2 + by_w^2 + cx_w + dy_w \tag{3}$$

where a, b, c, and d are coefficients obtained from fitting the measurement data from industry process shown in Figure 1 (b) [4], (x_w, y_w) is across-wafer location. We obtain the coefficients of the above across-wafer variation model by fitting the industrial 65nm process measured ring oscillator delay with 348 wafers from 23 lots. In this section, we assume that a, b, c, and d are fixed for a process. In practice, these coefficients may vary slightly from wafer-to-wafer or lot-to-lot. Figure 2 illustrates the PDF of the fitting coefficients for 348 wafers. From the figure, we find that the coefficients are distributed within 30% of the mean. The most accurate way is to model them as random variables. However, this will significantly increase the complexity of the variation model. For simplicity, in this paper, we assume the coefficients to be constant (using the mean value). Making such assumption introduces some error of the model, we will further discuss how to reduce the error in Section IV. In the rest of this section, all simulations are based-on this extracted model.

Combining Equation (2) and (3), we have:

$$v_p(x_w, y_w) = ax_w^2 + by_w^2 + cx_w + dy_w + v_{d-d} + v_r \tag{4}$$

In the rest of this section, we will analyze spatial variation based on the above model. Table I summarizes the mathematical notations used in this section. In the rest of this paper, we assume that inter-die random variation v_{d-d} and within-die random variation v_r are Gaussian random variables with zero mean (the nonzero mean can be lumped in to systematic offset v_c)[5].

[3]We assume that the fitting error is purely random, that is, it only introduces inter-die variation without affecting within-die variation. We further discuss the impact of fitting error in Section IV.

[4]Since we look on the wafer mean as wafer to wafer random variation and the systematic offset is lumped in to constant term v_c, there is no constant term in the quadratic across-wafer variation model (lumped to wafer to wafer variation and systematic offset).

[5]We assume inter-lot random, inter-wafer random, and inter-die random variation to be independent zero mean Gaussian random variables. Therefore, v_{d-d} is also a zero mean Gaussian random variable.

Symbols	Description	Units
	Across-wafer variation symbols	
V	variation source	1
v_c	constant systematic offset	1
v_0	nominal value	1
$v_p(x,y)$	Variation of within-die location (x,y)	1
v_{af}	Across-wafer variation (quadratic function)	1
v_{d-d}	Inter-die random variation (zero mean Gaussian)	1
v_r	Within-die random variation (zero mean Gaussian)	1
σ_{d-d}^2	Variance of v_{d-d}	1
σ_r^2	Variance of v_r	1
a, b	Across-wafer variation coefficients	mm^{-2}
c, d	Across-wafer variation coefficients	mm^{-1}
	Inter-die/spatial/within-die variation symbols	
v_g	Inter-die variation	1
v_s	Within-die spatial variation	1
v_l	Within-die random variation	1
	Size/location symbols	
r_w	Wafer radius	mm
(l_x, l_y)	x and y dimension die size	mm
(x_w, y_w)	Within-wafer location	mm
(x_c, y_c)	Location of the center of the die in the wafer	mm
(x, y)	Within-die location	mm
ω	Angle between the die and wafer coordinates	1
(x', y')	Within-die location in wafer coordinate $x' = x\cos\omega + y\sin\omega$, $y' = y\cos\omega - x\sin\omega$	mm
(l'_x, l'_y)	$l'_x = l_x\cos\omega + l_y\sin\omega$, $l'_y = l_y\cos\omega - l_x\sin\omega$	mm
(x'', y'')	$x'' = x'\sqrt{a/b} + c/(2\sqrt{ab})$, $y'' = y'\sqrt{b/a} + d/(2\sqrt{ab})$	mm
(l''_x, l''_y)	$l''_x = l'_x\sqrt{a/b} + c/(2\sqrt{ab})$, $l''_y = l'_y\sqrt{b/a} + d/(2\sqrt{ab})$	mm
$r_{d\mu}$	$r_{d\mu} = \sqrt{bx''^2 + ay''^2}$	1
$r_{d\sigma}$	$r_{d\sigma} = \sqrt{x''^2 + y''^2}$	mm
δ	Euclidean distance between (x''_1, y''_1) and (x''_2, y''_2) $\delta = \sqrt{(x''_1 - x''_2)^2 + (y''_1 - y''_2)^2}$	mm
r''_m	$r''_m = \sqrt{l''^2_x/4 + l''^2_y/4}$	mm
	Other symbols	
k_0	$k_0 = r_w^2(a+b)/4 - c^2/4a - d^2/4b$	1
k_1	$k_1 = r_w^4(a^2+b^2)/16 - r_w^4 ab/24 + \sigma_{d-d}^2 s$	1
k_2	$k_2 = k_1/(abr_w^2)$	mm^2
α	$\alpha = x''_1 x''_2 + y''_1 y''_2$	mm^2
β	$\beta = \sigma_r^2/(abr_w^2)$	mm^2
s_0	$s_0 = \cos^2\omega(al_x^2 + bl_y^2)/12 + \sin^2\omega(bl_x^2 + al_y^2)/12$	1
s_1	$s_1 = s_0 + c^2/4a + d^2/4b$	1

TABLE I: Notations. Note: Unit 1 means a no unit. In this paper, we assume that variation is normalized with respect to the nominal value, hence variation has no unit.

A. Variation of Mean and Variance with Location

Equation (4) provides a wafer level variation model, however, in real design, only die level variation model can be applied, i.e., for a die, whose center lies on (x_c, y_c) wafer coordinates, we want to know the variation of location (x, y) (assuming the coordinate of the center of the die to be $(0,0)$). In order to obtain the die level variation, we have to obtain the across-wafer coordinate from the die location in the wafer (x_c, y_c) and within-die location (x, y). In this paper, we assume that the chip coordinate aligns with the chip edges and the wafer coordinate aligns with the major and minor axises of the across-wafer variation parabola[6]. Notice that in practice, the wafer coordinate and chip coordinate might not be aligned, as shown in Figure 3, where ω is the angle between wafer coordinate and chip coordinate. We may convert die location (x, y) to wafer coordinate (x', y') by rotating coordinates, as shown in Table I. In this case, the within wafer location of within-die location (x, y) is calculated as:

$$x_w = x_c + x' \qquad y_2 = y_c + y'$$

[6]If we force the wafer coordinate and chip coordinate to be aligned, there will be a crossing term in the across-wafer variation model in Equation (4), which makes problem more complicated.

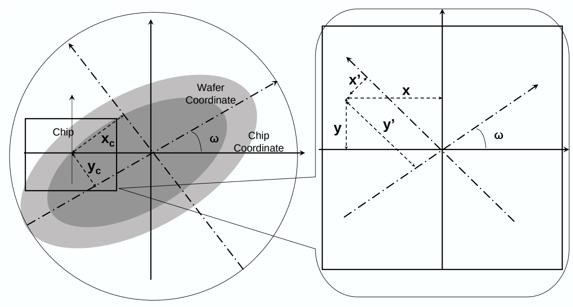

Fig. 3: Wafer coordinate and chip coordinate.

Then variation of location (x,y) is calculated as:

$$v_p(x,y) = a(x_c+x')^2+b(y_c+y')^2+ \\ c(x_c+x')+d(y_c+y')+v_{d-d}+v_r \quad (5)$$

In real design flow, the die location in the wafer (x_c,y_c) is not known to designers. We can convert the wafer-level systematic variation model to a die-level model by noting that dies are always distributed evenly in the wafer. Therefore, we may model (x_c,y_c) as random variables which are evenly distributed in the circle centering at $(0,0)$ with radius r_w (radius of the wafer). For simplicity, we convert rectangular coordinate to polar coordinate:

$$x = \rho\cos\theta \qquad y = \rho\sin\theta \quad (6)$$

where ρ and θ are independent random variables. ρ is with triangle distribution ranging from 0 to r_w, θ is with uniform distribution ranging from 0 to 2π:

$$PDF_\rho(\rho) = 2\rho/r_w^2 \qquad 0\le\rho<r_w \quad (7)$$
$$PDF_\theta(\theta) = 1/2\pi \qquad 0\le\theta<2\pi$$

With PDF, we can also obtain the first few order moments and joint moments of x_c and y_c. Since (x_c,y_c) are distributed in a symmetric area, joint moment $E[x_c^m y_c^n]=0$ when either m or n is odd number. Therefore, we only need to consider the even order moments and joint moments:

$$E[x_c^2]=E[y_c^2]=r_w^2/4 \quad (8)$$
$$E[x_c^4]=E[y_c^4]=r_w^4/8$$
$$E[x_c^2y_c^2]=r_w^4/24$$

The detailed derivation of the above equations is in Appendix A. In this case, the variation at location (x,y), $v_p(x,y)$, is expressed as a function of four random variables: x_c, y_c, v_{d-d}, and v_r. Then, the mean of $v_p(x,y)$ is calculated as:

$$\mu_{v_p}(x,y) = E[v_{aw}(x_c+x',y_c+y')]+E[v_{d-d}]+E[v_r] \quad (9)$$

As discussed above, v_{d-d} and v_r are zero mean, $v_{aw}(x,y)$ is quadratic function of x_c and y_c, therefore, $E[v_{aw}(x,y)]$ can be obtained from the moments and joint moments of x_c and y_c as shown in Equation (8):

$$\mu_{v_p}(x,y) = k_0 + r_{d\mu}^2 \quad (10)$$

where $r_{d\mu}$ and k_0 are defined in Table I. In a way similar to mean calculation, we may also calculate variance of $v_p(x,y)$:

$$\sigma_{v_p}^2(x,y) = k_1+\sigma_r^2+abr_w^2r_{d\sigma}^2 \quad (11)$$

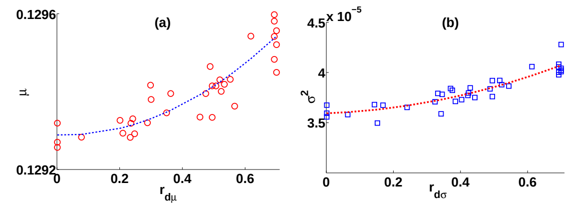

Fig. 4: (a) μ change for different $r_{d\mu}$, (b) σ^2 change for different $r_{d\sigma}$.

where k_1 and $r_{d\sigma}$ are defined in Table I.

The detailed derivation of Equation (10) and (11) is in Appendix B. From Equation (10) and (11), it is interesting to note that different within-die locations may have different means and variances [7]. The location (x_0,y_0) having the smallest mean and variance is given by letting $x''=0$ and $y''=0$:

$$x''=0 \quad\Rightarrow\quad x_0=-c\cos\omega/2a-d\sin\omega/2b$$
$$y''=0 \quad\Rightarrow\quad y_0=d\cos\omega/2b-c\sin\omega/2a$$

The locations farther away from (x_0,y_0) will have larger mean and variance. Figure 4 illustrates the mean and variance for different $r_{d\mu}$ (or $r_{d\sigma}$) obtained from our proposed model as shown in Equation (5). From the figure, we find that the mean and variance differ for different on chip locations, but the difference is very small. Especially for mean, the difference is less than 1%. Therefore, in the real measurement data, the location dependence of mean and variance is not obvious because a very small noise will overwhelm the difference.

B. Appearance of Spatial Correlation

Besides mean and variance, we are also interested in the covariance between two locations (x_1,y_1) and (x_2,y_2). Similar to the calculation of mean and variance, covariance is calculated as:

$$Cov = k_1+abr_w^2\alpha$$

Knowing the variance and covariance calculated above, we may obtain the correlation coefficient as:

$$\rho = \sqrt{\frac{k_2^2+2k_2\alpha+\alpha^2}{(k_2+\beta)^2+(r_{d\sigma1}^2+r_{d\sigma2}^2)(k_2+\beta)+r_{d\sigma1}^2r_{d\sigma2}^2}} \quad (12)$$

where α, β, and k_2 are defined in Table I. The detailed derivation of covariance and correlation coefficient is in Appendix C. From Equation (12), we obtain the upper bound and lower bound of the correlation coefficient for a certain Euclidean distance:

$$\rho\le\rho_u = \sqrt{1-\frac{\delta^2k_2+\delta^2\beta/2+2\beta k_2+\beta^2}{(k_2+\beta)^2+2r_m''^2(k_2+\beta)+r_m''^4}}$$

$$\rho\ge\rho_l = \sqrt{1-\frac{\delta^2(k_2-r_m''^2/2+\delta^2/4)+\beta(\beta+2k_2+2r_m''^2)+r_m''^4}{(k_2+\beta)^2+\delta^2(k_2+\beta)/2+\delta^4/16}}$$

[7] Such difference is caused by the chip-level nonlinearity of the across-wafer variation function (we assume quadratic function as in Equation(3)). If the the across-wafer variation function is linear at chip level, for example, a piecewise linear function with piece size larger than chip size, the mean and variance will be the same for all locations of a die.

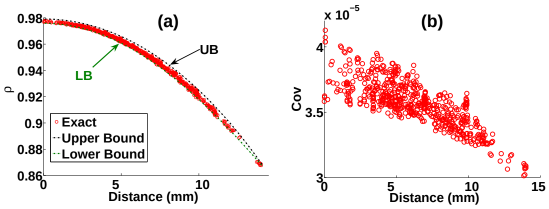

Fig. 5: (a) Apparent spatial correlation and (b) covariance as a function of distance.

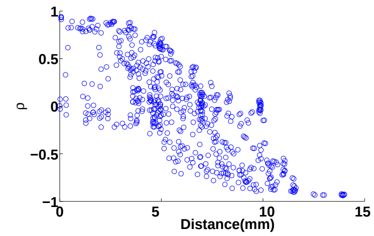

Fig. 6: Correlation cofficient for within-die spatial variation after inter-die variation is removed.

where δ, l''_x , l''_y, and r''_m are defined in Table I. From the upper bound and lower bound, we may also calculate the range of correlation coefficient:

$$\rho_u - \rho_l \le \sqrt{4r''^2_m/(r''^2_m + k_2 + \beta)}$$

The derivation of the upper bound, lower bound, and range of correlation coefficient is in Appendix D. Notice that usually the wafer size is much larger than the die size, that is $k_2 \gg r''^2_m$, therefore, $\rho_u - \rho_l \ll 1$, that is, the range of correlation coefficient for a certain distance is very narrow. Moreover, from the above equation, we also find that when the variances of the inter-die random and within-die random variation increase, the range decreases. This explains why most existing works [2], [9] find that spatial correlation is a function of distance.

Figure 5(a) illustrates the exact data for 40 locations, the upper bound and the lower bound obtained from our proposed model as shown in Equation (5). From the figure, we find that the range of ρ for a certain distance is very narrow. Although the correlation coefficient is within a narrow range, covariance is not, as shown in Figure 5(b). This is because of the differences of variance across the die.

Figure 6[8] shows the correlation coefficient for within-die variation after subtracting the mean variation of the die (mainly caused by across wafer variation). In the figure, the correlation coefficients are obtained from our proposed model as shown in Equation (5).

We observe that the within-die spatial variation is almost *NOT* spatially correlated, as empirically observed in [15], [17], [18]. This further validates that the spatial variation is caused by systematic across-wafer variation.

C. Dependence between Inter-die and Within-die Variation

In most existing variation models, process variation is decomposed into inter-die, within-die spatial, and within-die random variation:

$$v_p = v_g + v_s + v_l \tag{13}$$

where v_g is the inter-die variation, v_s is the within-die spatial variation, and v_l is the within-die variation. Usually v_g is modeled as the variation of the chip mean, v_s is the residual of across-wafer variation after subtracting the inter-die components, and v_l is the pure random local variation. v_g, v_s, and v_l are assumed to be independent.

With the variation model in Equation (5), we may also calculate the inter-die, within-die spatial, and within die random variation. Within-die random variation is the local random variation: $v_l = v_r$. Inter-die and spatial variation is induced by the die-to-die variation, and across-wafer variation. Inter die variation is calculated as the variation of the chip mean:

$$\begin{aligned} v_g &= \frac{1}{l_x l_y} \iint_{\substack{|x|<l_x/2 \\ |y|<l_y/2}} v_p(x,y)dxdy \\ &= ax_c^2 + by_c^2 + cx_c + dy_c + v_{d-d} + s_0 \end{aligned}$$

where s_0 is defined in Table I. Within-die spatial variation is calculated as the residual of across-wafer variation after subtracting the chip mean:

$$\begin{aligned} v_s(x,y) &= v_p(x,y) - v_g - v_l \\ &= r^2_{d\mu} + 2ax'x_c + 2by'y_c - s_1 \end{aligned} \tag{14}$$

where s_1 is defined in Table I. The derivation of the above equation is shown in Appendix E. From the above equations, we find that both inter-die and within-die spatial variations are functions of random variables x_c and y_c. Hence, we may not decompose process variation into independent inter-die and within-die spatial variation.

D. When can Spatial Variation be Ignored?

In this section, we analyze the accuracy of the simple two-level inter-/within-die variation model for different chip sizes. If we only consider inter-/within-die variation, we may lump the across-wafer variation into inter-die variation, that is, approximate the across-wafer variation as a piecewise constant function, as shown in Figure 7(a). To evaluate the impact of the approximation error, we may treat such approximation error as noise and the process variation as signal; and then evaluate the signal to noise ratio. In order to do this, we calculate the mean square approximation error and the total variance of variation. The signal to noise ratio when ignoring the spatial variation is given as:

$$\begin{aligned} SNR &= \sigma^2_{total}/MSE \\ &\approx \frac{6abr_w^4 + 6(c+d)r_w^2 + \sigma_M^2 + \sigma_R^2}{abr_w^2(l_x^2 + l_y^2) + 2(c+d)l_x l_y} \end{aligned}$$

It can be seen that MSE depends on chip size. When chip size is small, MSE is small. This is because we approximate

[8] In the figure, the correlation coefficient can be a negative number when distance is large. This is because after subtracting the mean, when the within-die variation of one corner increases, the within-die variation of the opposite corner must decrease. That means, the within-die variations of opposite corners are negative correlated. Moreover even when two locations are very closed, if they lie on the opposite side of the center, their correlation is still near zero.

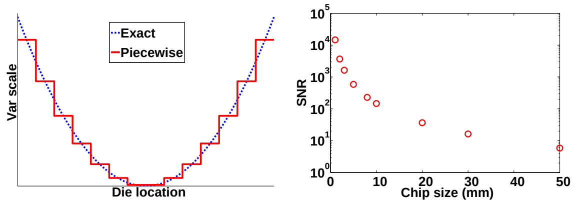

Fig. 7: Approximating across-wafer variation. Note: We assume square chips and chip size means edge length in *mm*. (a) Piecewise constant, (b) SNR V.S. chip size.

the across-wafer variation as a piecewise constant function with small steps, hence such approximation is accurate. Figure 7(b) illustrates the SNR for different die sizes. It can be seen that the SNR decreases when die size increases as expected. We also observe that when chip size (l_x and l_y) is smaller than 1cm, the SNR is up to 100. That means, two-level inter-/within-die variation model is accurate.

IV. General across-wafer Variation Model

In the previous section, we assumed that the across-wafer variation is a quadratic function as shown in Equation (5). In practice, across-wafer variation may not be an exact parabola. Moreover, the across-wafer variation will be slightly different for different wafers. Therefore, there will be some fitting residual after subtracting the across-wafer parabola:

$$v(x_w, y_w) = v_p(x_w, y_w) + v_e(x_w, y_w) \tag{15}$$

where v_p is the quadratic across-wafer variation model as shown in Equation (5) and v_e is fitting residual. In the previous section, we assume that the fitting residual is lumped into inter-die random variation. However, the fitting residual contains not only inter-die random variation but a systematic trend of within-die variation. Figure 8(a) illustrates the original delay variation across the wafer, and Figure 8(b) illustrates the fitting residual of a wafer delay variation after subtracting the quadratic across-wafer variation function. From the figure, we find that after removing the quadratic across-wafer variation, the scale of process variation reduces dramatically. From the die point of view, such residual will also introduce some spatial correlation. Figure 9(a) illustrates the correlation coefficients for different distances of the fitting residual. We find that the correlation coefficient is still high (≥ 0.4) the fitting residual, but the correlation coefficients are no longer within a narrow band for a given distance. From Figure 8(b), we also find the the spatial frequencies of the fitting residual are low. Therefore, from the die point of view, the fitting residual can be approximated by the first order Taylor expansion:

$$\begin{aligned} v_e(x', y') &= v_e(x_c + x', y_c + y') \approx v_e(x_c, y_c) + s_x x' + s_y y' \quad (16) \\ s_x &= \partial v_e(x_w, y_w)/\partial x_w | x_w = x_c, y_w = y_c \\ s_y &= \partial v_e(x_w, y_w)/\partial y_w | x_w = x_c, y_w = y_c \end{aligned}$$

where $v_e(x', y')$ is the fitting residual at die location (x', y'). In the above model, the term $v_e(x_c, y_c)$ is the same for the whole chip, but different from chip to chip. We may lump it into die-to-die variation. Since the impact of fitting residual on different chips is different, s_x and s_y vary from chip to chip. In this case, we may model s_x and s_y as random variables. Figure 10 illustrates the distribution of s_x and s_y obtained from measurement data of process 2. In order to obtain samples of s_x and s_y, we remove the quadratic wafer-level spatial pattern for each wafer, and then fit linear model in Equation 16 for each die. From the figure, we find that both s_x and s_y follow Gaussian distribution. Moreover, the correlation between s_x and s_y is very weak ($\rho < 0.1$). Therefore, in this paper, we assume that s_x and s_y are uncorrelated Gaussian random variables.

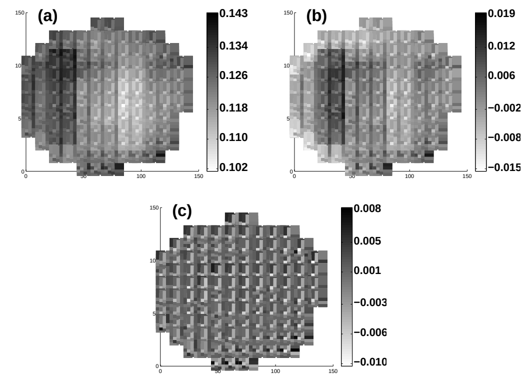

Fig. 8: Ring oscillator frequency within a wafer. (a) Original delay variation; (b)Residual after subtracting Equation(5); Residual after subtracting Equation (16).

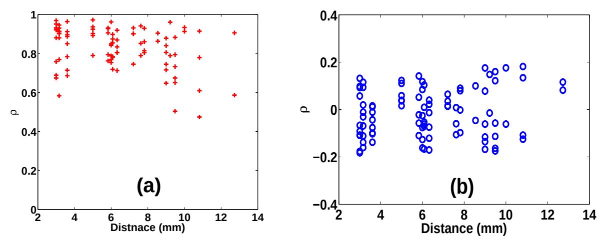

Fig. 9: (a) Correlation coefficient after subtracting Equation(5), (b) Correlation coefficient after subtracting Equation (16).

Notice that when we model the fitting residual as a linear within-die variation trend, two more random variables s_x and s_y are introduced. This makes the variation model more complicated. When the across-wafer variation is with a perfect parabola and the fitting residual is not significant, we may just lump the fitting residual in to inter-die and random variation.

In addition, we also observe that after subtracting the model of fitting residual in Equation (16), the remaining variation is almost uncorrelated, as illustrated in Figure 8(c) and Figure 9(b).

Combining Equation (16) and (5), we obtain a general die-level across-wafer variation model:

$$\begin{aligned} v_p(x, y) = {} & a(x_c + x')^2 + b(y_c + y')^2 + c(x_c + x') + \\ & d(y_c + y') + s_x x' + s_y y' + v_{d-d} + m_r(x, y) \end{aligned} \tag{17}$$

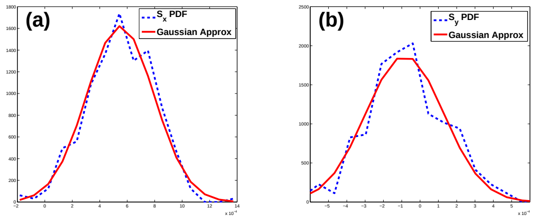

Fig. 10: (a) PDF of s_x, (b) PDF of s_y.

V. Modeling Spatial Variability

As discussed in Section I, spatial variation largely comes from the deterministic across-wafer variation. Hence, modeling the within-die variation as spatial-correlated random variables is not accurate as discussed in Section III.

In this section, we introduce three new spatial variation models considering across-wafer variation:

- *S*lope *A*ugmented *A*cross-*W*afer variation model (*SAAW*).
- *Q*uadratic *A*cross-*W*afer variation model (*QAW*).
- *L*ocation *D*ependent *A*cross-*W*afer variation model (*LDAW*).

In the rest of this section, we will discuss these models in detail:

A. Slope Augmented Across-Wafer Model

Equation (17) calculates the variation for a given location (x,y). In the equation, the die location within the wafer (x_c,y_c) are modeled as random variables and their PDF is shown in Equations (6) and (7). Equation (17) provides a new spatial variation model. We refer to the new model as *S*lope *A*ugmented *A*cross-*W*afer variation model (*SAAW*).

Notice that in *SAAW* model, there are only six random variables, inter-die random variation v_{d-d}, within-die random variation v_r, die location within the wafer x_c and y_c, and slope of fitting residual s_x and s_y. However, for the traditional distance-based spatial variation model, the number of spatial variation sources depends on the number of grids. Larger chip needs more variables. Therefore, our new model not only models the across-wafer variation accurately but also is more efficient than the traditional spatial correlation model.

B. Quadratic Across-Wafer Model

In our original variation model in [28], we model the across-wafer variation as a quadratic function, as shown in Equation (5), without modeling the fitting residual. In this case, there are only four random variables, x_c and y_c, v_{d-d}, and v_r. We refer to this model as *Q*uadratic *A*cross-*W*afer variation model (*QAW*). Since *QAW* does not consider fitting residual, it is not as accurate as *SAAW*. As discussed in Section IV, when the across-wafer variation is a perfect parabola, the fitting residual is not significant[9], we may just lump the fitting residual into inter-die random variation and simplify *SAAW* to *QAW*.

[9]For example, process 1 as discussed in Section II.

C. Location Dependent Across-Wafer Model

As discussed in Section III-D, when die size is small enough, applying the two-level inter-/within-die variation model does not introduce much error. However, inter-/within-die variation model still does not consider the mean and variance difference at different locations of a chip, as discussed in Section III-A. To further improve the accuracy of inter-/within-die variation model, we may account for this:

$$v(x,y) \approx v'_d + \mu_{v_p}(x,y) + \sigma_{v_p}(x,y)v'_r(x,y) \tag{18}$$

where v'_d is inter-die variation including inter-lot random, inter-wafer random, inter-die random, and across-wafer variation; $v'_r(x,y)$ is within die variation including within-die random variation and residual of across-wafer variation; $\mu_{v_p}(x,y)$ and $\sigma_{v_p}(x,y)$ are mean and variance difference at different locations of a chip, which can be calculated from Equation (10) and (11). We refer to the above model as *L*ocation *D*ependent *A*cross-*W*afer variation model (*LDAW*). *LDAW* is a further simplification of *QAW*, it lumps the across-wafer variation into inter-die and within-die variation. Inter-die variation is modeled as chip mean as discussed in Section III-C, and the residual is lumped into within-die variation. In this paper, we assume that v'_d is zero mean Gaussian random variable. The variance of v'_d is obtained from measurement. We also assume $v'_r(x,y)$ to have a standard normal distribution. In this case, the within-die variation, $\sigma_{v_p}(x,y)v'_r(x,y)$, is a zero mean Gaussian random variable whose variance is $\sigma^2_{v_p}(x,y)$, which is determined by within-die location. Moreover, in this model, the mean of $v(x,y)$ is $\mu_{v_p}(x,y)$ which is also location dependent. Notice that in Equation 18, $\mu_{v_p}(x,y)$ and $\sigma_{v_p}(x,y)$ are deterministic value for a certain within-die location (x,y). Therefore, *LDAW* model has only two random variables: v'_d and v'_r which is the same as two level inter-/within-die model. Hence compared to inter-/within-die model, *LDAW* has similar efficiency but higher accuracy because *LDAW* considers mean and variance difference across the chip while inter-/within-die model does not.

VI. Application to Statistical Timing Analysis

In this section, we apply our across-wafer variation model to statistical static timing analysis.

A. Delay Model

In statistical timing analysis, people usually approximate cell delay as linear function of variation sources:

$$D = D_0 + A_s V^T \tag{19}$$

where D_0 is the nominal cell delay, $A_s = (a_{s1}, a_{s2}, \ldots, a_{sn})$ is the vector of linear sensitivity coefficients, and $V = (V_1, V_2, \ldots, V_n)$ is the vector of variation sources. For *SAAW* and *QAW*, since each variation source is a quadratic function of random variables, as show in Equation (17) and (5), the gate delay is a second order function of random variables:

$$D = p_2(RV)$$

where $p_i(\cdot)$ is a i^{th} order polynomial function and $RV = (rv_1, rv_2, \ldots, rv_n)$ is the vector of random variables (such as

x_c and y_c). In this case, the quadratic SSTA flow in [29], [30] can be applied to estimate chip delay variation. For *LDAW*, each variation source is a linear function of random variables, as shown in Equation (18), then the cell delay is also a linear function of random variables:

$$D = p_1(RV)$$

In this case, the linear SSTA flow in [29], [30] can be applied to estimate chip delay variation.

To model cell delay more accurately, a quadratic cell delay model [29]–[32] can be used :

$$D = D_0 + A_s V^T + V B_s V^T \tag{20}$$

where $B_s = (b_{sij})$ is matrix of second order sensitivity coefficients. In this case, for *LDAW*, the cell delay is a quadratic function of random variables. Therefore, quadratic SSTA can be applied to estimate chip delay. However, for *SAAW* and *QAW*, since each variation source is a quadratic function of random variables, the cell delay becomes a 4^{th} order function of random variables:

$$D = p_4(RV)$$

Handling such high order delay variation function is complicated. In this case, the moment matching technique in [29], [30] is applied to approximate the 4^{th} order function to a quadratic function by matching the first two order moments and joint moments:

$$\begin{aligned}
p_2(RV) &\approx p_4(RV) \\
E[p_2(RV)] &= [p_4(RV)] \\
E[rv_i \cdot p_2(RV)] &= E[rv_i \cdot p_4(RV)] \\
E[rv_i^2 \cdot p_2(RV)] &= E[rv_i^2 \cdot p_4(RV)] \\
E[rv_i rv_j \cdot p_2(RV)] &= E[rv_i rv_j \cdot p_4(RV)]
\end{aligned}$$

With the above approximation, quadratic SSTA can be applied. Notice that moment matching approximation is performed only once for all cells and does not increase the run time of SSTA.

Moreover, it was shown in [29], [30] that ignoring crossing terms of quadratic cell delay function (*semi-quadratic* delay model) significantly improves the run time of SSTA without affecting accuracy too much. Therefore, to improve efficiency, we may ignore the crossing terms when we perform quadratic SSTA (*semi-quadratic* SSTA).

B. Experimental Result

We have implemented the non-linear SSTA in [29], [30] with different spatial variation model in C++. In order to verify the efficiency and accuracy, three comparison cases are defined: 1) Monte-Carlo simulation with the exact deterministic across-wafer variation model[10], which is the golden case for comparison; 2) distance-based spatial correlation model from [2], which is referred to as *SP*atial *C*orrelation model (*SPC*); 3) two-level inter-/within die variation model, which is referred to as *I*nter-/*W*ithin-die variation model (*IW*).

We apply all the above methods to the ISCAS85 suite of benchmarks in *P*redictive *T*echnology *M*odel (*PTM*) 45nm technology [33]. We assume random placement for ISCAS85 circuits. Since process variation has smaller impact on interconnect delay than on logic cell delay, we only consider logic cell delay when calculate the full chip delay variation. In the experiment, we consider the gate length variation obtained from minimum square error fitting on the ring oscillator delay from industrial 65nm process (Process 2 as discussed in Section II) measurement from the model as shown in Equation (17). We obtain the across-wafer coefficients a, b, c, and d, fitting residual residual coefficients s_x and s_y, standard deviation of random inter-wafer, inter-die, and within-die variation as percentage with respect to the nominal value. Then we assume that the percentages of all the above coefficients to nominal value are the same at 45nm technology node and 65nm technology node. To obtain across-wafer coefficients a, b, c, and d, we apply quadratic function to fit the across-wafer variation for each wafer to obtain the fitting coefficients for each wafer, then use average coefficients of all wafers for our experiment. To obtain fitting residual residual coefficients s_x and s_y, We first get the slope of fitting residual s_x and s_y for each chip, and then calculate the mean and variance of s_x and s_y for all chips. In the experiment, we assume s_x and s_y to be Gaussian random variables with mean and variance obtained from the measurement data.

1) Full Chip Delay: In the experiment, we assume that the chips size is 2cm×2cm and the wafer radius is 15cm. Since ISCAS85 benchmarks are very small, the impact of spatial variation on delay is not significant within the circuit. In order to show such impact,we assume the benchmarks are stretched on a 2cm×2cm chip. In our experiment, for the *SPC* model, we divide the chip to $10 \times 10 = 100$ grids. Table II illustrates the percentage error of mean (μ), standard deviation (σ), and 95-percentile point (95%) and run time (T) of different variation models. In the table, we also compare the result of using quadratic cell delay model (Quad) and linear cell delay model (Lin). We only use quadratic cell delay model for golden case simulation (exact), the error is calculated as error of different variation models compared to the golden case simulation. For *SAAW* and *QAW*, we also compare the results of applying quadratic SSTA with crossing terms (*SAAW* Quad and *QAW* Quad) and applying SSTA without crossing terms (*SAAW* S-Quad and *QAW* S-Quad). From the table, we have the following observations:

- Compared to full quadratic SSTA, *semi-quadratic* SSTA (SSTA without crossing terms) achieves up to 8X speed up with less than 1% accuracy loss.
- *SAAW* is more accurate than *QAW*. This is because the fitting residual is significant for the measurement data, *QAW* ignores fitting residual and hence introduces more error.
- The error of *SAAW* using *semi-quadratic* SSTA is within 2% while the error of spatial correlation model is up to 6.5%.
- Compared to quadratic cell delay model, linear cell delay

[10] In the simulation, each wafer may have different across-wafer variation which is obtained from measurement data of process 2. We have simulated 318 wafers correspondent to 318 measured wafers.

Bench-mark	Chip size	Exact			*SAAW*			*LDAW*			*SPC*			*IW*		
		μ	σ	95%	μ	σ	95%	μ	σ	95%	μ	σ	95%	μ	σ	95%
c1908	10	17.4	2.24	24.2	-1.2	-1.8	-1.9	-2.2	-5.5	-5.9	-1.7	-3.5	-3.3	-2.5	-6.8	-7.1
	6	17.5	2.23	24.3	-1.1	-1.5	-1.6	-2.2	-4.1	-4.2	-1.2	-2.6	-2.4	-2.2	-4.5	-4.3
	3	17.6	2.22	24.4	-0.9	-1.2	-1.1	-1.1	-1.8	-1.6	-1.0	-1.5	-1.5	-1.9	-2.1	-2.5
c3540	10	25.6	3.42	34.6	-1.0	-2.2	-1.6	-1.4	-6.2	-4.9	-1.8	-5.4	-4.8	-2.2	-7.5	-6.9
	6	25.8	3.45	34.3	-0.8	-1.2	-1.0	-1.2	-3.8	-3.3	-1.5	-3.4	-3.0	-1.3	-4.2	-3.7
	3	25.8	3.44	34.4	-0.5	-1.0	-1.0	-1.1	-2.1	-2.0	-1.0	-1.9	-1.9	-1.1	-2.2	-2.3
c7552	10	48.9	6.47	64.7	-1.2	-1.1	-1.4	-2.8	-3.5	-3.7	-2.5	-2.2	-2.7	-3.2	-5.6	-6.3
	6	48.9	6.47	64.7	-1.0	-1.1	-1.3	-1.4	-2.5	-2.3	-2.2	-2.1	-2.5	-2.9	-3.1	-3.3
	3	48.9	6.47	64.7	-0.6	-0.9	-1.0	-1.0	-1.1	-1.2	-0.9	-1.3	-1.4	-1.3	-1.4	-1.6

TABLE III: Percentage error for ISCAS85 benchmark stretching on a chip with different chip size. Note: We assume square chips and chip size means edge length in *mm*. The exact delay values are in *ns*.

has less than 2% accuracy loss. This is because in our experiment, the cell delay variation is well approximated by a linear function.

- For linear cell delay model, *SAAW* achieves about 6X speed up compared to *SPC*. This is because there are 100 grids in the spatial correlation model, resulting in 37 spatial random variables[11], while *SAAW* has only 6 random variables.
- *LDAW* and *IW* are very efficient. However, both models have much larger error than others. This is because both model ignore correlation. But *LDAW* is still more accurate than *IW* without run time penalty.

Since linear cell delay model and semi-quadratic SSTA are accurate, we assume linear cell delay and apply semi-quadratic SSTA for all experiments. Moreover, since *SAAW* is more accurate than *QAW* with only a small run time overhead, we do not consider the *QAW* model in the following experiments.

In our experiment, we only consider big chips. As discussed in Section III-D, when the chip size is small, the impact on across-wafer variation at die level is not significant. In order to verify this, we perform delay estimation of ISCAS85 benchmarks stretching on different size chips. Table III shows the percentage error for different models with different chip size. From the table, we find that when chip size is small, *LDAW* and *IW* is accurate. Considering that *LDAW* has similar run time but is more accurate (although when chip size is small, the accuracy improvement is limited) compared to *IW*, *LDAW* is always better than *IW*.

2) Delay of Blocks on Different Locations on a Chip: The above experiment assumes that the benchmarks are stretched on a chip. However, in real design, especially for big chips, the design is separated into several blocks and each block only occupies a small region on a chip. In this case, the critical path is within a small region instead of spanning all over the chip. As discussed in Section III-A, different chip locations may have different mean and variance. Therefore, when a block is place at different locations of a chip, its delay variation may be different. In order to show such effect, we assume that the ISCAS85 benchmark circuit is placed (no stretched) in different locations of a chip: center (C), lower left corner (LL), lower right corner (LR), upper left corner (UL), and upper right corner (UR), and then calculate the delay variation with location. Since ISCAS85 benchmarks are very small, the impact of spatial variation on delay is not significant within the circuit. Therefore, in this experiment, we only compare two models *LDAW* and *IW*[12].

Table IV compares the percentage error of *LDAW* and *IW* for ISCAS85 benchmarks placed on different locations of a 2cm×2cm chip. From the table, we find that the error of *LDAW* is within 1% error from the exact simulation and the error of *IW* is up to 8%. This is because *LDAW* predicts different mean and variance for different location correctly, as discussed in Section V while *IW* can only give the same mean and variance for all locations.

bench mark	loca-tion	Exact			*LDAW*			*IW*		
		μ	σ	95%	μ	σ	95%	μ	σ	95%
c3540	C	25.4	3.29	32.2	+0.4	+0.3	+0.3	+1.1	+2.4	+2.8
	LL	24.8	3.22	31.9	+0.8	+0.6	+0.3	+3.5	+1.4	+1.9
	LR	26.2	3.35	33.1	-0.7	+0.6	-0.6	-1.9	-2.4	-1.8
	UL	26.5	3.36	33.3	-0.2	+0.3	+0.3	-3.3	-3.0	-4.4
	UR	27.1	3.41	34.1	-0.3	-0.3	-0.6	-6.1	-4.1	-5.2
c7552	C	48.2	6.37	60.2	+0.8	+0.3	+0.2	+1.0	+0.9	+0.9
	LL	47.2	6.11	58.5	+0.4	+0.3	+0.7	+3.6	+4.6	+3.1
	LR	49.4	6.51	62.3	-0.2	+0.3	+0.3	-0.8	-1.9	-3.0
	UL	49.5	6.65	63.1	-0.2	+0.1	+0.1	-1.0	-4.0	-4.1
	UR	50.1	6.91	65.3	-0.4	+0.3	+0.1	-1.0	-7.4	-7.4

TABLE IV: Delay percentage error at different locations in a $2cm \times 2cm$ chip. Note: The exact delay values are in *ns*.

Table V, VI, and VII shows percentage error of *LDAW* and *IW* for ISCAS85 benchmarks placed on different locations of a 1cm×1cm, 6mm×6mm, and 3mm×3mm chip, respectively. From the tables, we find that the error of *IW* becomes smaller when chip size is small.

bench mark	loca-tion	Exact			*LDAW*			*IW*		
		μ	σ	95%	μ	σ	95%	μ	σ	95%
c3540	C	25.4	3.26	31.9	+0.5	+0.4	+0.2	+1.3	+1.6	+1.9
	LL	25.0	3.25	31.6	+0.5	+0.3	+0.3	+2.4	+1.3	+2.4
	LR	25.8	3.30	32.4	-0.4	-0.4	-0.4	-1.1	-0.9	-0.8
	UL	26.1	3.32	32.6	-0.4	+0.3	-0.2	-2.0	-1.5	-1.4
	UR	26.2	3.34	33.0	-0.3	-0.3	-0.6	-2.6	-1.8	-2.2
c7552	C	48.6	6.38	60.4	+0.2	-0.1	-0.2	+1.0	+0.5	+0.6
	LL	48.2	6.35	60.3	-0.2	-0.2	+0.2	+1.7	+1.0	+1.1
	LR	48.9	6.42	60.4	+0.2	-0.2	+0.2	-0.2	-0.7	-0.5
	UL	49.1	6.43	61.0	-0.2	-0.2	-0.2	-0.9	+1.0	-1.3
	UR	49.2	6.45	61.2	-0.3	-0.4	-0.5	-1.1	-1.4	-1.8

TABLE V: Delay comparison for ISCAS85 benchmark in $1cm \times 1cm$ chip. Note: The exact delay values are in *ns*.

bench mark	loca-tion	Exact			*LDAW*			*IW*		
		μ	σ	95%	μ	σ	95%	μ	σ	95%
c3540	C	25.5	3.27	32.0	+0.4	+0.2	+0.3	+0.9	+0.7	+1.4
	LL	25.1	3.25	31.7	+0.5	+0.3	+0.3	+2.4	+1.3	+2.4
	LR	26.0	3.32	32.6	-0.4	-0.4	-0.4	-1.1	-0.9	-0.8
	UL	26.2	3.34	32.8	-0.4	+0.3	-0.2	-2.0	-1.5	-1.4
	UR	26.3	3.35	33.1	-0.3	-0.3	-0.6	-2.6	-1.8	-2.2
c7552	C	48.4	6.37	60.1	+0.2	-0.1	-0.2	+1.0	+0.5	+0.6
	LL	48.1	6.34	59.9	-0.2	-0.2	+0.2	+1.7	+1.0	+1.1
	LR	49.0	6.43	60.7	+0.2	-0.2	+0.2	-0.2	-0.7	-0.5
	UL	49.3	6.46	61.3	-0.2	-0.2	-0.2	-0.9	+1.0	-1.3
	UR	49.4	6.48	61.5	-0.3	-0.4	-0.5	-1.1	-1.4	-1.8

TABLE VI: Delay comparison for ISCAS85 benchmark in $6mm \times 6mm$ chip. Note: The exact delay values are in *ns*.

[11] There are 100 correlated spatial random variables, we apply PCA to truncate some insignificant principle components and there remains 37 significant principle components.

[12] When the circuit is in a small region, *SAAW* and *QAW* will give similar result as *LDAW*, and *SPC* will give similar result as *IW*.

Bench-mark	delay model	Exact			*SAAW* Quad				*SAAW* S-Quad				*QAW* Quad				*QAW* S-Quad				*LDAW**				*SPC**				*IW**			
		μ	σ	95%	μ	σ	95%	T	μ	σ	95%	T	μ	σ	95%	T	μ	σ	95%	T	μ	σ	95%	T	μ	σ	95%	T	μ	σ	95%	T
c1908	Quad	17.6	2.27	24.5	-0.4	-0.9	-1.0	146	-0.9	-1.7	-1.7	27	-0.8	-1.9	-2.3	54	-1.3	-2.5	-3.2	19	-2.1	-7.5	-6.9	9	-1.5	-4.2	-3.8	1450	-2.6	-10.2	-8.9	8
	Lin	-	-	-	-0.8	-1.5	-1.4	150	-1.4	-1.8	-2.0	26	-0.9	-3.6	-3.1	53	-1.2	-3.4	-3.9	18	-2.6	-7.5	-8.1	10	-2.1	-4.4	-4.0	135	-3.0	-11.5	-10.3	10
c3540	Quad	25.7	3.43	34.5	+0.4	+0.9	+0.7	212	-0.4	-1.3	-1.1	36	+0.4	-1.8	-1.2	76	-0.9	-2.1	-1.9	25	+0.4	-5.8	-4.6	13	-1.2	-4.8	-4.0	4210	-1.4	-7.3	-6.5	12
	Lin	-	-	-	-0.6	-1.2	-1.2	209	-1.1	-1.9	-1.6	35	-0.9	-3.6	-3.1	77	-1.2	-5.1	-3.9	27	-1.8	-6.5	-6.0	9	-2.0	-6.5	-5.7	202	-2.9	-9.3	-8.8	10
c7552	Quad	48.9	6.47	64.7	-0.6	+0.3	+0.2	435	-0.8	-0.2	-0.9	67	-0.8	-1.6	-1.4	115	-1.6	-1.5	-1.7	48	-2.7	-3.6	-4.0	20	-1.0	-2.3	-2.9	8182	-2.1	-6.7	-6.5	22
	Lin	-	-	-	-0.6	-0.5	-0.6	430	-1.5	-1.4	-1.6	101	-1.1	-1.3	-1.4	109	-1.9	-2.2	-2.8	79	-3.3	-4.6	-4.9	16	-3.3	-3.5	-4.3	433	-3.9	-8.9	-8.2	15

TABLE II: Delay percentage error for different variation models. Note: the μ, σ, and 95-percentile point for exact simulation is in *ns*. Run time (T) is in *ms*. * for *LDAW*, *SPC*, and *IW*, linear SSTA is applied when assuming linear cell delay model.

bench mark	loca-tion	Exact			*LDAW*			*IW*		
		μ	σ	95%	μ	σ	95%	μ	σ	95%
c3540	C	25.6	3.28	32.2	+0.4	+0.2	+0.3	+0.5	+0.3	+0.8
	LL	25.4	3.26	320	+0.5	+0.6	+0.3	+1.2	+1.0	+1.2
	LR	25.8	3.31	32.4	-0.2	-0.4	-0.6	-0.5	-0.7	-0.7
	UL	25.9	3.22	32.5	-0.2	+0.3	-0.2	-1.0	-1.1	-0.9
	UR	26.0	3.31	32.7	-0.2	-0.2	-0.3	-0.7	-0.8	-1.1
c7552	C	48.6	6.38	60.3	+0.2	-0.3	+0.7	+0.2	+0.5	+1.1
	LL	48.4	6.37	60.1	-0.2	0.1	+0.2	+0.4	+0.3	+0.7
	LR	48.8	6.42	60.6	+0.2	-0.3	+0.3	-0.6	-0.9	-0.5
	UL	48.9	6.43	60.9	-0.2	+0.2	-0.2	-0.4	+0.0	-0.6
	UR	49.2	6.45	61.2	-0.2	-0.2	-0.3	-0.7	-0.8	-1.1

TABLE VII: Delay comparison for ISCAS85 benchmark in $3mm \times 3mm$ chip. Note: The exact delay values are in *ns*.

VII. Application to Statistical Leakage Analysis

Besides SSTA, we also apply our variation model to statistical leakage power analysis. Usually, cell leakage power variation is modeled as exponential function of variation sources:

$$P_{leak} = P_0 \cdot e^{\sum c_i V_i} \tag{21}$$

where P_0 is the nominal leakage power and c_i's are sensitivity coefficients. The full chip leakage power is calculated as the sum of leakage power of all cells:

$$P_{chip} = \sum_{i \in Cell} P_{i,leak} \tag{22}$$

where $Cell$ is the set of all cells in the chip and $P_{i,leak}$ is leakage power of the i^{th} cell. Since each variation source is a quadratic function as in Equation (17), the cell leakage power is an exponential of a quadratic function of random variables. Considering the random variables may be non-Gaussian, there is no closed-form equations to calculate the full chip leakage power. Therefore, in this paper, we apply Monte-Carlo simulation to obtain the full chip leakage power variation.

We have implemented leakage variation analysis with different models in Matlab. In the experiment, we use the same setting and comparison cases as the SSTA experiment in Section VI. For each variation model, we use 100,000 sample Monte-Carlo simulation to obtain the full chip leakage power for all variation models. For the leakage analysis, we assume that 900 copies of ISCAS benchmark circuits are placed in a 30×30 array on a 2cm×2cm chip. Table VIII compares the leakage variation for ISCAS85 benchmarks. From the table, we observe that:

- Error of *SAAW* is within 5% while error of *SPC* is up to 17%. Moreover, *SAAW* is 7X faster than *SPC* because there are fewer random variables for *SAAW*.
- *SAAW* is more accurate than *QAW*, but is about 50% slower.
- Both *LDAW* and *IW* are not accurate. This is because both these models does not consider correlation and hence under estimate the leakage power variation.

Similar to the SSTA, for leakage variation analysis, we also perform leakage estimation in different size chips: 1cm×1cm, 6mm×6mm, and 3mm×3mm. For the 1cm×1cm chip, we assume that 225 copies of ISCAS85 benchmark circuits are placed in a 15×15 array, for the 6mm×6mm chip, we assume 100 copies of ISCAS85 benchmark circuits are placed in a 10×10 array, and for the 3mm×3mm chip, we assume 25 copies of ISCAS benchmark circuits are placed in a 5×5 array. Table IX shows the error percentage for different models for different size chips. From the table, we find that the error of *LDAW*, *SPC*, and *IW* reduces when chip size becomes smaller as expected.

Bench-mark	Chip size	Exact			*SAAW*			*LDAW*			*SPC*			*IW*		
		μ	σ	95%	μ	σ	95%	μ	σ	95%	μ	σ	95%	μ	σ	95%
c1355	10	15.4	3.52	23.2	+1.2	+3.1	+3.0	-4.7	-10.6	-11.2	+4.8	+7.5	+9.2	-5.4	-12.3	-13.8
	6	5.92	1.62	10.3	+1.0	+2.7	+2.9	-3.5	-7.5	-8.3	+2.7	+6.2	+6.7	-3.9	-8.1	-9.2
	3	1.48	0.40	2.58	+0.6	+1.8	+2.0	-1.8	-3.5	-3.7	+1.6	+2.9	+3.1	-2.0	-3.5	-4.0
c1908	10	23.9	5.7	36.1	+1.0	+2.7	+2.9	-4.2	-12.3	-14.1	+3.7	+8.5	+9.2	-5.5	-13.9	-15.1
	6	10.6	2.25	16.1	+0.9	+2.0	+2.1	-3.4	-7.3	-8.2	+1.9	+5.6	+6.8	-3.5	-7.3	-9.0
	3	2.65	0.57	4.03	+1.0	+2.2	+2.2	-1.3	-3.5	-4.0	+1.2	+2.9	+3.1	-1.9	-4.0	-4.4
c2670	10	32.8	5.72	45.2	+1.2	+2.8	+2.9	-5.3	-11.1	-14.0	+4.2	+8.2	+9.1	-6.5	-12.3	-17.1
	6	14.6	2.55	20.1	+1.0	+1.8	+1.9	-3.5	-6.2	-7.1	+2.4	+4.3	+6.0	-4.3	-7.1	-8.3
	3	3.65	0.65	5.03	+0.8	+1.4	+1.7	-1.9	-3.2	-3.7	+2.0	+3.6	+3.5	-2.2	-4.5	-5.1
c3540	10	50.5	9.37	71.1	+1.2	+1.8	+2.1	-3.9	-7.8	-10.5	+2.9	+5.3	+6.2	-5.0	-9.6	-14.5
	6	22.3	4.16	30.2	+1.0	+1.4	+1.7	-2.3	-4.5	-5.5	+1.8	+3.5	+3.6	-3.4	-6.1	-8.5
	3	5.58	1.05	7.55	+0.9	+1.2	+1.4	-1.2	-3.1	-3.0	+1.0	+1.4	+2.0	-1.4	-3.4	-3.9
c7552	10	102	18.5	141	+1.4	+2.3	+2.4	-4.6	-7.3	-9.9	+4.0	+6.2	+8.6	-6.3	-8.2	-12.5
	6	45.0	8.19	6.26	+1.2	+1.9	+1.9	-3.0	-5.2	-7.3	+2.7	+4.2	+5.1	-3.5	-6.0	-8.2
	3	11.4	2.06	1.58	+0.9	+0.7	+1.3	-1.2	-1.8	-2.0	+1.5	+1.6	+1.6	-2.3	-2.9	-3.4

TABLE IX: Leakage error for different variation model on different size chips. Note: exact values are in *mW*.

VIII. Summary of Different Models

In the previous sections, we compared the accuracy and efficiency of different models. Table X summarizes the advantages and disadvantages of our proposed spatial variation models (*SAAW*, *QAW*, and *LDAW*), and the traditional variation models (*SPC* and *IW*). Our proposed across-wafer variation models exactly model the across-wafer variation and the number of random variables does not depend on chip size. Therefore they are accurate and efficient. *SAAW* has six random variables and it can be applied to any across-wafer variation models. *QAW* has four random variables, hence it is more efficient than *SAAW*. However, it can be applied only when the across-wafer variation is a perfect parabola. *LDAW* is the most efficient, ignores correlation and only works for small chips. Moreover,

Bench-	Exact			*SAAW*				*QAW*				*LDAW*				*SPC*				*IW*			
mark	μ	σ	95%	μ	σ	95%	T	μ	σ	95%	T	μ	σ	95%	T	μ	σ	95%	T	μ	σ	95%	T
c1355	62.1	14.5	92.5	+1.5	+3.3	+3.2	16.3	+3.4	+6.5	+5.4	9.8	-5.5	-15.6	-16.9	2.5	+5.3	+10.4	+12.2	123	-7.9	-19.6	-20.9	2.7
c1908	95.6	20.3	144	+0.9	+4.4	+3.7	15.5	+2.3	+7.8	+6.3	9.7	-6.5	-17.8	-19.6	2.6	+5.7	+14.8	+14.3	122	-8.6	-19.2	-23.5	2.9
c2670	131	22.9	181	+1.4	+2.7	+1.7	15.7	+2.9	+8.5	+4.7	9.5	-7.9	-16.9	-22.0	2.8	+6.8	+12.2	+9.4	122	-9.2	-20.3	-25.5	2.4
c3540	201	37.4	282	+1.5	+2.3	+1.8	15.4	+3.1	+5.8	+4.4	10.4	-5.6	-16.5	-20.2	2.6	+4.9	+11.2	+8.2	123	-8.3	-18.2	-23.5	2.7
c7552	403	73.2	562	+1.6	+2.7	+1.9	15.3	+3.7	+6.0	+5.0	10.1	-7.3	-12.6	-16.9	2.6	+7.1	+13.8	+10.7	122	-9.2	-20.3	-24.5	2.5

TABLE VIII: Leakage error percentage for different models in $2cm \times 2cm$ chip. Note: The exact values are in mW. Run time (T) is in s.

SAAW and *QAW* need to know the across-wafer variation. Therefore, one needs to track the die locations within the wafer to build up the model.

On the other hand, the traditional variation models as well as *LDAW*) only require measurement on a die without tracking die locations. Therefore, they are somewhat easier to build. However, such models are not accurate compared to our proposed models. Moreover, for *SPC*, since the number of random variables depends on number of grids, it is not as efficient as our proposed models.

IX. Arbitrary Across-Wafer Variation Function

In this paper, we assume across-wafer variation to be a parabola. In most cases, our proposed *SAAW* model is good enough to model the across-wafer variation at die level. However, there are some special cases where the across wafer variation is an arbitrary function as follows:

$$v_p = f(x_w, y_w) + v_{d-d} + v_r$$

In this case, the statistical characteristics such as mean, variance, covariance, and correlation coefficient depend on the function f. In most of the cases, we may not have the closed form formulae to calculate the statistical characteristics. However, we may still apply similar method as in Section III to model the across-wafer location for a die as random variables:

$$v_p(x,y) = f(x_c + x, y_c + y) + v_{d-d} + v_r$$

When we know the function f, either in closed form or as a numerical lookup table, we may perform Monte-Carlo simulation on the above formula for statistical analysis. In this case, since there is no closed-form, we cannot perform analytical statistical analysis, such as SSTA or statistical leakage analysis.

X. Conclusion

In this paper, we analytically study the impact of systematic across-wafer variation on within-die spatial variation. For simplicity, we first assume that across-wafer variation is a quadratic function. We first observe that different locations on a chip may have different means and variances and such difference becomes more significant when chip size increases. Secondly, we find that spatial correlation is visible only when the across wafer systematic is not taken into account. When it is taken into account, we show that within die random variability does not exhibit a strong or useful pattern of spatial correlation. We exploited these observations in order to create a much more accurate and efficient model for performance variability prediction. Thirdly, we find that the within-die spatial variation is *NOT* independent of the inter-die variation. However, when chip size is small enough, such dependence is weak and the across-wafer variation can be lumped in to inter-die variation. In this case, the two level inter-/within-die variation model is still accurate. We further consider the case when the across-wafer variation is not with a perfect quadratic function. Based on the above analysis, we have proposed accurate and efficient variation models for deterministic across wafer variation. We further apply our new variation models to two applications: statistical static timing analysis and statistical leakage analysis. Experimental result shows that compared to the distance-based spatial variation model, our new model reduces the error from 6.5% to 2% for statistical timing analysis and reduces error from 17% to 5% for statistical leakage analysis. Our model also improves the run time by 6X for statistical timing analysis and by 7X for statistical leakage analysis.

Appendix

A. Moments of x_c and y_c

2^{nd} and 4^{th} order moments of x_c and y_c:

$$\begin{aligned}
E[x_c^2] &= E[\rho^2]E[\cos^2\theta] = \int_0^{r_w} \rho^2 \cdot 2\rho/r_w^2 d\rho \int_0^{2\pi} \cos^2(\theta)/2\pi d\theta \\
&= (\rho^4/2r_w^2)|_0^{r_w} \cdot (\theta + \sin(\theta)\cos(\theta))|_0^{2\pi}/4\pi \\
&= r_w^2/4 \\
E[x_c^4] &= E[\rho^4]E[\cos^4\theta] = \int_0^{r_w} \rho^4 \cdot 2\rho/r_w^2 d\rho \int_0^{2\pi} \cos^4(\theta)/2\pi d\theta \\
&= (\rho^6/3r_w^2)|_0^{r_w} \cdot (12\theta + 8\sin(2\theta) + \sin(4\theta))|_0^{2\pi}/64\pi \\
&= r_w^4/8
\end{aligned}$$

Since x_c and y_c are symmetric, we have:

$$\begin{aligned}
E[y_c^2] &= E[x_c^2] = r_w^2/4 \\
E[y_c^4] &= E[x_c^4] = r_w^4/8 \qquad \square
\end{aligned}$$

Joint moment of x_c and y_c:

$$\begin{aligned}
E[x_c^2 y_c^2] &= E[\rho^4]E[\cos^2\theta \sin^2\theta] \\
&= \int_0^{r_w} \rho^4 \cdot 2\rho/r_w^2 d\rho \int_0^{2\pi} \cos^2(\theta)\sin^2(\theta)/2\pi d\theta \\
&= (\rho^6/3r_w^2)|_0^{r_w} \cdot (4\theta - \sin(4\theta))|_0^{2\pi}/64\pi \\
&= r_w^4/24 \qquad \square
\end{aligned}$$

B. Mean and variance of v_p

We first express v_{aw} as:

$$\begin{aligned}
v_{aw}(x_c + x', y_c + y') &= a(x_c + x')^2 + b(y_c + y')^2 + c(x_c + x') + d(y_c + y') \\
&= a(x_c + x''\sqrt{b/a})^2 + b(y_c + y''\sqrt{a/b})^2 - c^2/4a - d^2/4b
\end{aligned}$$

Model Type	Advantages	Disadvantages	Models	# of RVs	Case to Apply
Across-wafer models	Accurate Efficient	Need die tracking to extract	*SAAW* Equ(17)	6	large chip, non-parabola across-wafer variation
			QAW Equ (5)	4	large chip, parabola across-wafer variation
			LDAW Equ (18)	2	small chip
Traditional models	Easy to extract	Not accurate	*SPC*	Depend on # of grids	large chip
			IW	2	small chip

TABLE X: Summary of different variation models.

Mean of v_p: We first compute the mean of v_{aw} as follows, notice that we only need to consider even order moments and joint moments of x_c and y_c as discussed in Section III:

$$\begin{aligned}\mu_{v_{aw}} &= E[v_{aw}(x_c+x',y_c+y')]\\ &= E[a(x_c+x''\sqrt{b/a})^2+b(y_c+y''\sqrt{a/b})^2]-c^2/4a-d^2/4b\\ &= aE[x_c^2]+bE[y_c^2]+bx''^2+ay''^2-c^2/4a-d^2/4b\\ &= r_w^2(a+b)/4-c^2/4a-d^2/4b+bx''^2+ay''^2\\ &= k_0+r_{d\mu}^2\end{aligned}$$

Since we assume that v_{d-d} and v_r are with zero mean, mean of v_p is:

$$\mu_{v_p}=\mu_{aw}+\mu_{d-d}+\mu_r=k_0+r_{d\mu}^2 \qquad \square$$

Variance of v_p: We first compute the variance of v_{aw}:

$$\begin{aligned}\sigma_{v_{aw}}^2 &= E[v_{aw}^2(x_c+x',y_c+y')]-E^2[v_{aw}(x_c+x',y_c+y')]\\ &= a^2E[x_c^4]+4abx''^2E[x_c^2]+b^2E[y_c^4]+4aby''^2E[y_c^2]+2abE[x_c^2y_c^2]-(aE[x_c^2]+bE[y_c^2])^2\\ &= r_w^4(a^2+b^2)/16-r_w^4ab/24+abr_w^2(x''^2+y''^2)\end{aligned}$$

Then the variance of v_p is:

$$\begin{aligned}\sigma_{v_p}^2 &= \sigma_{v_{aw}}^2+\sigma_{v_{d-d}}^2+\sigma_{v_r}^2\\ &= r_w^4(a^2+b^2)/16-r_w^4ab/24+\sigma_{d-d}^2+\sigma_r^2+abr_w^2(x''^2+y''^2)\\ &= k_1+\sigma_r^2+abr_w^2r_{d\sigma}^2 \qquad \square\end{aligned}$$

C. Covariance and correlation coefficient between $v_p(x_1,y_1)$ and $v_p(x_2,y_2)$

Covariance between $v_p(x_1,y_1)$ and $v_p(x_2,y_2)$: We fist compute the covariance between $v_{aw}(x_1,y_1)$ and $v_{aw}(x_2,y_2)$:

$$\begin{aligned}\mathrm{cov}_{aw} &= E[v_{aw}(x_c+x_1',y_c+y_1')\cdot v_{aw}(x_c+x_2',y_c+y_2')]-\\ &\quad E[v_{aw}(x_c+x_1',y_c+y_1')]\cdot E[v_{aw}(x_c+x_2',y_c+y_2')]\\ &= a^2E[x_c^4]+4abx_1''y_2''E[x_c^2]+b^2E[y_c^4]+4abx_2''y_1''E[y_c^2]+\\ &\quad 2abE[x_c^2y_c^2]-(aE[x_c^2]+bE[y_c^2])^2\\ &= r_w^4(a^2+b^2)/16-r_w^4ab/24+abr_w^2(x_1''y_2''+x_2''y_1'')\end{aligned}$$

Since all devices on the same chip share the same inter-die variation and within-die random variation is independent for different devices, then the covariance between $v_p(x_1,y_1)$ and $v_p(x_2,y_2)$ is:

$$\begin{aligned}\mathrm{cov} &= \mathrm{cov}_{aw}+\sigma_{d-d}^2\\ &= r_w^4(a^2+b^2)/16-r_w^4ab/24+\sigma_{d-d}^2+abr_w^2(x_1''y_2''+x_2''y_1'')\\ &= k_1+abr_w^2\alpha \qquad \square\end{aligned}$$

Correlation coefficient $v_p(x_1,y_1)$ and $v_p(x_2,y_2)$:

$$\begin{aligned}\rho &= \frac{\mathrm{cov}}{\sigma_1\sigma_2}=\sqrt{\frac{\mathrm{cov}^2}{\sigma_1^2\sigma_2^2}}\\ &= \sqrt{\frac{(k_1+abr_w^2\alpha)^2}{(k_1+\sigma_r^2+abr_w^2r_{d\sigma1}^2)(k_1+\sigma_r^2+abr_w^2r_{d\sigma2}^2)}}\\ &= \sqrt{\frac{(k_1/(abr_w^2)+\alpha)^2}{(k_1/(abr_w^2)+\sigma_r^2/(abr_w^2)+r_{d\sigma1}^2)(k_1/(abr_w^2)+\sigma_r^2(abr_w^2)+r_{d\sigma2}^2)}}\\ &= \sqrt{\frac{(k_2+\alpha)^2}{(k_2+\beta+r_{d\sigma1}^2)(k_2+\beta+r_{d\sigma2}^2)}}\\ &= \sqrt{\frac{k_2^2+2k_2\alpha+\alpha^2}{(k_2+\beta)^2+(r_{d\sigma1}^2+r_{d\sigma2}^2)(k_2+\beta)+r_{d\sigma1}^2r_{d\sigma2}^2}} \qquad \square\end{aligned}$$

D. Upper bound and lower bound of ρ

Upper bound of ρ:

$$\begin{aligned}\rho &= \sqrt{\frac{k_2^2+2k_2\alpha+\alpha^2}{(k_2+\beta)^2+(r_{d\sigma1}^2+r_{d\sigma2}^2)(k_2+\beta)+r_{d\sigma1}^2r_{d\sigma2}^2}}\\ &= \sqrt{1-\frac{2k_2\beta+\beta^2+\delta^2k_2+(x_1''y_2''-x_2''y_1'')^2+(r_{d\sigma1}^2+r_{d\sigma1}^2)\beta}{(k_2+\beta)^2+(r_{d\sigma1}^2+r_{d\sigma2}^2)(k_2+\beta)+r_{d\sigma1}^2r_{d\sigma2}^2}}\end{aligned} \tag{23}$$

In the above equation, ρ is represented in a form of: $\sqrt{1-\zeta/\eta}$. To obtain the upper bound, we increase the denominator η and reduce numerator ζ. Considering that:

$$\begin{aligned}(x_1''y_2''-x_2''y_1'')^2 &\geq 0\\ r_{d\sigma1}^2+r_{d\sigma2}^2 &= x_1''^2+y_1''^2+x_2''^2+y_2''^2\\ &\geq ((x_1''-x_2'')^2+(y_1''-y_2'')^2)/2\\ &= \delta^2/2\end{aligned}$$

$$\begin{aligned}& -l_x''/2\leq x''\leq l_x'' \quad -l_y''/2\leq y''\leq l_y''\\ \Rightarrow\ & r_{d\sigma}^2=x''^2+y''^2\leq l_x''^2/4+l_y''^2/4=r_m''^2\\ \Rightarrow\ & r_{d\sigma1}^2+r_{d\sigma2}^2\leq 2r_m''^2 \quad r_{d\sigma1}^2r_{d\sigma2}^2\leq r_m''^4\end{aligned}$$

Replacing $(x_1''y_2''-x_2''y_1'')^2$ with 0 and $(r_{d\sigma1}^2+r_{d\sigma2}^2)$ with $\delta^2/2$ in the numerator, and replacing $(r_{d\sigma1}^2+r_{d\sigma2}^2)$ with $2r_m''^2$ and $r_{d\sigma1}^2r_{d\sigma2}^2$ with $r_m''^4$ in the denominator, we have:

$$\rho\leq\sqrt{1-\frac{\delta^2k_2+\delta^2\beta/2+2\beta k_2+\beta^2}{(k_2+\beta)^2+2r_m''^2(k_2+\beta)+r_m''^4}} \qquad \square$$

Lower bound of ρ: ρ is represented in a form of $\sqrt{1-\zeta/\eta}$ as shown in Equation (23). Consider that ζ/η is between 0 and 1, increasing ζ and η with the same value will increase ζ/η, then reduces ρ. Therefore, to obtain the lower bound, we first add a non-negative value $(r_{d\sigma1}^2-r_{d\sigma1}^2)^2/4$ to both numerator and denominator:

$$\begin{aligned}\rho &\geq \sqrt{1-\frac{2k_2\beta+\beta^2+\delta^2k_2+(x_1''y_2''-x_2''y_1'')^2+(r_{d\sigma1}^2+r_{d\sigma1}^2)\beta+(r_{d\sigma1}^2-r_{d\sigma1}^2)^2/4}{(k_2+\beta)^2+(r_{d\sigma1}^2+r_{d\sigma2}^2)(k_2+\beta)+r_{d\sigma1}^2r_{d\sigma2}^2+(r_{d\sigma1}^2-r_{d\sigma1}^2)^2/4}}\\ &= \sqrt{1-\frac{2k_2\beta+\beta^2+\delta^2k_2+(x_1''y_2''-x_2''y_1'')^2+(r_{d\sigma1}^2+r_{d\sigma1}^2)\beta+(r_{d\sigma1}^2-r_{d\sigma1}^2)^2/4}{(k_2+\beta)^2+(r_{d\sigma1}^2+r_{d\sigma2}^2)(k_2+\beta)+(r_{d\sigma1}^2+r_{d\sigma1}^2)^2/4}}\\ &= \sqrt{1-\frac{2k_2\beta+\beta^2+\delta^2k_2+(r_{d\sigma1}^2+r_{d\sigma1}^2-\delta^2/2)^2/4+3\delta^4/16+(r_{d\sigma1}^2+r_{d\sigma1}^2)\beta}{(k_2+\beta)^2+(r_{d\sigma1}^2+r_{d\sigma2}^2)(k_2+\beta)+(r_{d\sigma1}^2+r_{d\sigma1}^2)/4}}\end{aligned}$$

Considering that:

$$2r_m''^2\geq r_{d\sigma1}^2+r_{d\sigma2}^2\geq\delta^2/2\Rightarrow(2r_m''^2-\delta^2/2)^2\geq(r_{d\sigma1}^2+r_{d\sigma2}^2-\delta^2/2)^2$$

Similar to the upper bound proof, replacing $(r_{d\sigma1}^2+r_{d\sigma2}^2-\delta^2/2)^2$ with $(2r_m''^2-\delta^2/2)^2$ and $(r_{d\sigma1}^2+r_{d\sigma2}^2)$ with $2r_m''^2$ in the numerator, and replacing $(r_{d\sigma1}^2+r_{d\sigma2}^2)$ with $\delta^2/2$ in the denominator, we have:

$$\begin{aligned}\rho &\geq \sqrt{1-\frac{2k_2\beta+\beta^2+\delta^2k_2+(2r_m''^2-\delta^2/2)^2/4+3\delta^4/16+2r_m''^2\beta}{(k_2+\beta)^2+\delta^2(k_2+\beta)/2+\delta^4/16}}\\ &= \sqrt{1-\frac{\delta^2(k_2-r_m''^2/2+\delta^2/4)+\beta(\beta+2k_2+2r_m''^2)+r_m''^4}{(k_2+\beta)^2+\delta^2(k_2+\beta)/2+\delta^4/16}} \qquad \square\end{aligned}$$

Range of ρ: To obtain the range of ρ, similar to the proof of lower bound, we add a non-negative value $2r_m''^2(k_2+\beta)/2+$

$r_m''^4 - \delta^2(k_2+\beta)/2 - \delta^4/16$ to both numerator and denominator of the lower bound to obtain a looser lower bound:

$$
\begin{aligned}
\rho_l \geq & \rho_l' \\
= & \sqrt{1 - \frac{\delta^2(k_2/2 - r_m''^2/2 + 3\delta^2/16 - \beta/2) + \beta(\beta + 2k_2 + 2r_m''^2) + 2r_m''^4 + 2r_m''^2(k_2+\beta)/2}{(k_2+\beta)^2 + 2r_m''^2(k_2+\beta) + r_m''^4}}
\end{aligned}
$$

Then, the range of ρ can be calculated as:

$$
\begin{aligned}
\rho_u - \rho_l = & \sqrt{(\rho_u - \rho_l)^2} \leq \sqrt{(\rho_u - \rho_l)(\rho_u + \rho_l)} = \sqrt{\rho_u^2 - \rho_l^2} \\
\leq & \sqrt{\rho_u^2 - \rho_l'^2} = \sqrt{\frac{k_2(2r_m''^2 - \delta^2/2) + \beta(4r_m''^2 - \delta^2) + 3\delta^4/16 + 2r_m''^4 - \delta^2 r_m''^2/2}{(k_2+\beta+r_m''^2)^2}} \\
\leq & \sqrt{\frac{2k_2 r_m''^2 + 4\beta r_m''^2 + 5r_m''^4}{(k_2+\beta+r_m''^2)^2}}
\end{aligned}
$$

Since usually the wafer size is much larger than the die size, that is $k_2 \gg r_m''^2$, then $2k_2 r_m''^2 \gg r_m''^4$. Therefore, we have:

$$
\rho_u - \rho_l \leq \sqrt{\frac{2k_2 r_m''^2 + 4\beta r_m''^2 + 2k_2 r_m''^2 + 4r_m''^4}{(k_2+\beta+r_m''^2)^2}} = \sqrt{\frac{4r_m''^2}{k_2+\beta+r_m''^2}} \qquad \square
$$

E. Computation of v_g and v_s

Computation of v_g:

$$
\begin{aligned}
v_g = & \frac{1}{l_x l_y} \iint_{\substack{|x|<l_x/2 \\ |y|<l_y/2}} v_p(x,y)\,dxdy \\
= & ax_c^2 + by_c^2 + cx_c + dy_c + v_{d-d} + \\
& \frac{1}{l_x l_y} \iint_{\substack{|x|<l_x/2 \\ |y|<l_y/2}} \big(2(ax_c\cos\omega - by_c\sin\omega)x + 2(ax_c\sin\omega + by_c\cos\omega)y + (a\cos^2\omega + b\sin^2\omega)x^2 + \\
& (a\sin^2\omega + b\cos^2\omega)x^2 + 2(a-b)\cos\omega\sin\omega xy + v_r(x,y)\big)\,dxdy
\end{aligned}
$$

Since the integration region is symmetric, the integration of odd order moments and joint moments of x and y is zero. Moreover,notice that $v_r(x,y)$ is zero mean. Then, we have:

$$
\begin{aligned}
v_g = & ax_c^2 + by_c^2 + cx_c + dy_c + v_{d-d} + ((a\cos^2\omega + b\sin^2\omega)l_x^2 + (a\sin^2\omega + b\cos^2\omega)l_y^2)/12 \\
= & ax_c^2 + by_c^2 + cx_c + dy_c + v_{d-d} + \cos^2\omega(al_x^2 + bl_y^2)/12 + \sin^2\omega(bl_x^2 + al_y^2)/12 \\
= & ax_c^2 + by_c^2 + cx_c + dy_c + v_{d-d} + s_0 \qquad \square
\end{aligned}
$$

Computation of v_s:

$$
\begin{aligned}
v_s = & v_{aw}(x,y) + v_{d-d} - v_g \\
= & a(x_c + x')^2 + b(y_c + y')^2 + c(x_c + x') + d(y_c + y') + v_{d-d} - \\
& (ax_c^2 + by_c^2 + cx_c + dy_c + v_{d-d} + s_0) \\
= & 2ax'x_c + 2by'y_c + ax'^2 + by'^2 + cx' + dy' - s_0 \\
= & 2ax'x_c + 2by'y_c + bx''^2 + ay''^2 - s_0 - c^2/4a - d^2/4b \\
= & r_{d\mu}^2 + 2ax'x_c + 2by'y_c - s_1 \qquad \square
\end{aligned}
$$

REFERENCES

[1] H. Chang and S. S. Sapatnekar, "Statistical timing analysis considering spatial correlations using a single pert-like traversal," in *Proc. Int. Conf. on Computer Aided Design*, 2003.

[2] J. Xiong, V. Zolotov, and L. He, "Robust extraction of spatial correlation," in *Proc. Int. Symp. on Physical Design*, Apr 2006.

[3] D. Boning, J. Chung, D. Ouma, and R. Divecha, "Spatial variation in semiconductor processes: Modeling for control," in *Proceedings Process Control Diagnostics and Modeling in Semiconductor Manufacturing II Electrochemical Society Meeting*, May 1997.

[4] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," in *Proc. Design Automation Conf*, June 2003.

[5] F. Liu, "A general framework for spatial correlation modeling in vlsi design," in *Proc. Design Automation Conf*, Jun 2007.

[6] J.-H. Liu, M.-F. Tsai, L. Chen, and C. C.-P. Chen, "Accurate and analytical statistical spatial correlation modeling for vlsi dfm applications," in *Proc. Design Automation Conf*, Jun 2008.

[7] T. Sato, H. Ueyama, N. Nakayama, and K. Masu, "Determination of optimal polynomial regression function to decompose on-die systematic and random variations," in *Proc. Asia South Pacific Design Automation Conf.*, Feb 2008.

[8] A. Agarwal, D. Blaauw, V. Zolotov, S. Sundareswaran, M. Zhao, K. Gala, and R. Panda, "Statistical delay computation considering spatial correlations," in *Proc. Asia South Pacific Design Automation Conf.*, 2003.

[9] K. Chopra, N. Shenoy, and D. Blaauw, "Variogram based robust extraction of process variation," in *ACM/IEEE International Workshop on Timing Issues*, Feb 2007.

[10] W. Zhang, "An efficient method for chip-level statistical capacitance extractionion via random field simulation and production considering process variations with spatial correlation," in *Proc. Design Automation and Test in Europe*, Mar 2008.

[11] Q. Fu, W.-S. Luk, J. Tao, C. Yan, and X. Zeng, "Characterizing intra-die spatial correlation using spectral density method," in *International Symposium on Quality of Electronic Design*, Mar 2008.

[12] S. Ohkawa, H. Masuda, and Y. Inoue, "A novel expression of spatial correlation by a random curved surface model and its application to LSI designs," *IEICE Trans. on Fundamentals*, 2008.

[13] B. Liu, "Spatial correlation extraction via random field simulation and production chip performance regression," in *Proc. Design Automation and Test in Europe*, Mar 2008.

[14] W. Zhao, Y. Cao, F. Liu, K. Agarwal, D. Acharyya, S. Nassif, and K. Nowka, "Rigorous extraction of process variations for 65nm cmos design," in *Solid State Device Research Conference*, Sep 2007.

[15] P. Friedberg, W. Cheung, and C. J. Spanos, "Spatial modeling of micron-scale gate length variation," in *Proc. SPIE*, 2006.

[16] N. Drego, A. Chandrakasan, and D. Boning, "Lack of spatial correlation in mosfet threshold voltage variation and implications for voltage scaling," *IEEE Trans. on Semiconductor Manufacturing*, vol. 22, Feb 2009.

[17] Q. Zhang, C. Tang, J. Cain, A. Hui, T. Hsieh, N. Maccrae, B. Singh, K. Poolla, and C. J. Spanos, "Across-wafer cd uniformity control through lithography and etch process: experimental verification," in *Proc. SPIE*, 2007.

[18] K. Qian and C. J. Spanos, "A comprehensive model of process variability for statistical timing optimization," in *Proc. SPIE*, 2008.

[19] J. R. Sheats, *Microlithography Science and Technology*. 1998.

[20] J. Sali, S. Patil, S. Jadkar, and M. Takwale, "Hot-wire cvd growth simulation for thickness uniformity," in *International Conference on Cat-CVD Process*, 2001.

[21] S. Sakai, M. Ogino, R. Shimizu, and Y. Shimogaki, "Deposition uniformity control in a commercial scale HTO-CVD reactor," *Materials Research Society Symposium Proceedings*, 2007.

[22] J. Brcka and R. L. Robison, "Wafer redeposition impact on etch rate uniformity in ipvd system," *IEEE Trans.s on Plasma Sciences*, Feb 2007.

[23] T. W. Kim and E. S.Aydil, "Investigation of etch rate uniformity of 60 mhz plasma etching equipment," *Japanese Journal of Applied Physics*, Nov 2001.

[24] T. W. Kim and E. S.Aydil, "Effects of chamber wall conditions on cl concertration and si etch rate uniformity in palsma etching reactors," *Journal of The Electrochemical Society*, June 2003.

[25] Q. Zhang, K. Poola, and C. Spanos, "One step forward from run-to-run critical dimension control: Across-wafer level critical dimension control through lithography and etch process," *Journal of Process Control*, vol. 18, no. 10, pp. 937 – 945, 2008.

[26] B. Stine, D. Boning, and J. Chung, "Analysis and decomposition of spatial variation in integrated circuit processes and devices," in *IEEE Trans.s on Semiconductor Manufacturing*, pp. 24 – 41, Feb. 1997.

[27] K. Qian and C. J. Spanos, "Hierachical modeling of spatial variability of a 45nm test chip," in *SPIE*, 2009.

[28] L. Cheng, P. Gupta, C. Spanos, K. qian, and L. He, "Physically justifiable die-level modeling of spatial variation in view of systematic across wafer variability," in *Proc. Design Automation Conf*, 2009.

[29] L. Cheng, J. Xiong, and L. He, "Non-gaussian statistical timing analysis using Second-Order polynomial fitting," in *Proc. Asia South Pacific Design Automation Conf.*, Feb 2008.

[30] L. Cheng, J. Xiong, and L. He, "Non-gaussian statistical timing analysis using Second-Order polynomial fitting," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, Jan 2009.

[31] Y. Zhan, A. J. Strojwas, X. Li, and L. T. Pileggi, "Correlation-aware statistical timing analysis with non-gaussian delay distribution," in *Proc. Design Automation Conf*, pp. 77–82, June 2005. Anaheim, CA.

[32] L. Zhang, W. Chen, Y. Hu, J. A. Gubner, and C. C.-P. Chen, "Correlation-preserved non-gaussian statistical timing analysis with quadratic timing model," in *Proc. Design Automation Conf*, pp. 83 – 88, June 2005.

[33] Predictive Technology Model in *http://www.eas.asu.edu/ ptm/*.

Lerong Cheng received the B.S. degree in electronics and communication engineering from Zhongshan University, Guangzhou, China in 2001, the M.S. degree in Electrical and Computer Engineering from Portland State University in 2003, and the Ph.D degree in Electrical Engineering From University of California, Los Angeles in 2009. He is currently a CAD Engineer in SanDisk Corperation.

His research interests include computer-aided design of VLSI circuits and systems, programmable fabrics, low-power and high-performance designs, and statistical timing analysis.

Puneet Gupta (IEEE student member 2003-2007, member 2007-present, http://www.ee.ucla.edu/˜puneet) is currently a faculty member of the Electrical Engineering Department at UCLA. He received the B.Tech degree in Electrical Engineering from Indian Institute of Technology, Delhi in 2000 and Ph.D. in 2007 from University of California, San Diego. He co-founded Blaze DFM Inc. (acquired by Tela Inc.) in 2004 and served as its product architect till 2007.

He has authored over 60 papers, ten U.S. patents, and a book chapter. He is a recipient of NSF CAREER award, ACM/SIGDA Outstanding New Faculty Award, European Design Automation Association Outstanding Dissertation Award and IBM Ph.D. fellowship. Dr. Puneet Gupta has given tutorial talks at DAC, ICCAD, Intl. VLSI Design Conference and SPIE Advanced Lithography Symposium. He has served on the Technical Program Committee of DAC, ICCAD, ASPDAC, ISQED, ICCD, SLIP and VLSI Design. He served as the Program Chair of IEEE DFM&Y Workshop 2009, 2010.

Dr. Gupta's research has focused on building high-value bridges across application-architecture-implementation-fabrication interfaces for lowered cost and power, increased yield and improved predictability of integrated circuits and systems.

Costas J. Spanos received the Electrical Engineering Diploma from the National Technical University of Athens, Greece in 1980 and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from Carnegie Mellon University in 1981 and 1985, respectively. From 1985 to 1988 he was with the advanced Computer-Aided Design group of Digital Equipment Corporation, where he worked on the statistical characterization, simulation and diagnosis of VLSI processes. In 1988 he joined the Faculty at the department of Electrical Engineering and Computer Sciences of the University of California at Berkeley, where he is now a Professor. He was the Director of the Berkeley Microfabrication Laboratory from 1994 to 2000, the Director of the Electronics Research Laboratory from 2004 to 2005, and the Associate Dean for Research in the College of Engineering from 2004 to 2008. He is presently the Department Chair of EECS at UC Berkeley. Professor Spanos has served in the technical committees of numerous conferences and was the editor of the IEEE Transactions on Semiconductor Manufacturing from 1991 to 1994. He has published more than 200 referred articles, has received several best paper awards and has co-authored a textbook in semiconductor manufacturing. His present research interests include the application of statistical analysis in the design and fabrication of integrated circuits, and the development and deployment of novel sensors and computer-aided techniques in semiconductor manufacturing. He is also working towards the deployment of information technology and statistical data mining techniques for energy efficiency applications. In 2000 he was elected Fellow of the Institute of Electrical and Electronic Engineers for contributions and leadership in semiconductor manufacturing, and in 2009 he was appointed in the Andrew S. Grove Distinguished Professorship, in the Department of EECS.

Kun Qian (IEEE student member 2006 - present) received the B.S. degree from the department of Microelectronics at Peking University, Beijing, China in 2005. From 2005, He started his graduate study at the department of Electrical Engineering and Computer Sciences of the University of California at Berkeley, and is currently working towards the Ph.D. degree.

His research interests include characterization, analysis and modeling of semiconductor device and circuit variability, statistical compact model extraction, and yield predictions.

Lei He (IEEE M'99-SM'08) is a professor at electrical engineering department, University of California, Los Angeles (UCLA) and was a faculty member at University of Wisconsin, Madison between 1999 and 2002. He also held visiting or consulting positions with Cadence, Empyrean Soft, Hewlett-Package, Intel, and Synopsys, and was technical advisory board member for Apache Design Solutions and Rio Design Automation.

Dr. He obtained Ph.D. degree in computer science from UCLA in 1999. His research interests include modeling and simulation, VLSI circuits and systems, and cyber physical systems. He has published one book and over 200 technical papers with 12 best paper nominations mainly from Design Automation Conference and International Conference on Computer-Aided Design and five best paper or best contribution awards including the ACM Transactions on Electronic System Design Automation 2010 Best Paper Award.