Robust Extraction of Spatial Correlation

Jinjun Xiong, *Member, IEEE*, Vladimir Zolotov, *Senior Member, IEEE*, and Lei He, *Member, IEEE*

***Abstract*—The increased variability of process parameters makes it important yet challenging to extract the statistical characteristics and spatial correlation of process variation. Recent progress in statistical static-timing analysis also makes the extraction important for modern chip designs. Existing approaches extract either only a deterministic component of spatial variation or these approaches do not consider the actual difficulties in computing a valid spatial-correlation function, ignoring the fact that not every function and matrix can be used to describe the spatial correlation. Applying mathematical theories from random fields and convex analysis, we develop: 1) a robust technique to extract a valid spatial-correlation function by solving a constrained nonlinear optimization problem and 2) a robust technique to extract a valid spatial-correlation matrix by employing a modified alternative-projection algorithm. Our novel techniques guarantee to extract a valid spatial-correlation function and matrix from measurement data, even if those measurements are affected by unavoidable random noises. Experiment results, obtained from data generated by a Monte Carlo model, confirm the accuracy and robustness of our techniques and show that we are able to recover the correlation function and matrix with very high accuracy even in the presence of significant random noises.**

***Index Terms*—Extraction, modeling, nearest correlation matrix, process variation, spatial correlation, valid spatial correlation function.**

I. Introduction

AGGRESSIVE scaling down of transistors and interconnects has resulted in miraculous achievements in chip performance and functionality. This deep scaling of semiconductor technology, however, has introduced the problem of uncontrollable process variations. That is, we are unable to make transistors and interconnects with accurately predictable characteristics, let alone to make transistors the same on different copies of the same chip and even at different locations of the same chip. Thus, the only way to cope with variability is to take it into account during chip designs in order to maximize manufacturing yield. This consensus is supported by the development of statistical-static-timing-analysis (SSTA) tools capable of predicting statistical timing yield [1]–[3] of designed chips.

Modern SSTA tools can handle both interchip and intrachip random variations of process and environmental parameters. The interchip variations represent global variations that are the same for all devices on a given chip. The intrachip variations represent variations of devices within the same chip. The intrachip variations include spatially correlated variations and purely independent or uncorrelated variations. Spatial correlation describes the phenomenon that devices close to each other are more likely to have similar characteristics than devices far apart.

It is of importance to characterize process variation because that information is essential for any attempts to analyze or optimize designs statistically. For example, it is necessary to know the variations of device parameters in order to build the statistical delay models for both devices and interconnects, which are the inputs for both SSTA and robust circuit tuning. Recent SSTA techniques considering spatially correlated parameters [1], [4], [5], however, assume that the required spatial-correlation information given as a correlation matrix is known *a priori* and is always valid, i.e., the spatial-correlation matrix is always positive semidefinite. In fact, the only way to obtain these variation characteristics is to extract them from silicon measurements. Because of unavoidable measurement errors, there is no guarantee that the so-obtained correlation coefficients can form a valid correlation matrix.

To the best of our knowledge, no existing work has provided a detailed technique to extract that information properly from measurements except some preliminary exploration in [6]–[8]. The extraction of the deterministic component of L_{eff} variation was considered in detail in [9] for the 0.18-μm CMOS technology. But that publication ignored the random component of spatial variations, justifying its approach by the fact that for the 0.18-μm CMOS-technology random variations were not significant. Another two recent publications [10], [11] limited their consideration by simple computation of the spatial-correlation coefficient that is a function of distance, which is either a linear [10] or piecewise linear function [11]. There is no verification, however, that the extracted correlation function was a valid-correlation function, i.e., any correlation matrix generated from this function must be positive semidefinite. In fact, theoretically, as we will show in this paper, neither linear nor piecewise linear functions are valid spatial-correlation functions.

The major contribution of this paper is as follows. We provide the theoretical foundations for extracting the valid spatial-correlation information from silicon measurements. We develop a robust technique to extract a valid spatial-correlation function by solving a constrained nonlinear optimization problem. We also develop a robust technique to extract a valid spatial-correlation matrix by employing a modified alternative-projection algorithm. Our techniques are based upon the mathematical theories of random fields and convex analysis, and it is guaranteed that the resulting correlation function and correlation matrix are not only valid but also the closest ones

Manuscript received March 10, 2006; revised June 8, 2006. The work of J. Xiong was supported in part by the NSF under CAREER Award CCR-0401682 and in part by a UC MICRO Grant sponsored by Actel and Intel. A preliminary version of this paper was presented at ISPD 2006. This paper was recommended by Guest Editor J. Hu.

J. Xiong and V. Zolotov are with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: jinjun@us.ibm.com).

L. He is with the Electrical Engineering Department, University of California, Los Angeles, CA 90096 USA.

Digital Object Identifier 10.1109/TCAD.2006.884403

to the underlying model even if the data are distorted by significant measurement noises. Experiment results based upon a Monte Carlo model confirm the accuracy and robustness of our techniques. We achieve less than 10% errors for the extracted process variations, even if the measurement noise is more than 100% of the total process variations. Because of the promising results, we plan to apply our techniques to real wafer data to extract the spatial-correlation information in the future.

The rest of this paper is organized as follows. We first describe how to model process variations in Section II, then present our problem formulations in Section III, and provide algorithms to solve the problems in Sections IV and V, respectively. Experiment results are presented in Section VI, and we draw conclusions in Section VII.

II. Preliminary

A. Process-Variation Classification

There are two orthogonal ways to classify process variations. The first one is to classify the variations according to the scope of their occurrence [13]–[15] as follows: 1) Die-to-die (D2D) variation, which is also called interdie variation or between-die variation, describes the variation that affect parameters in different dies differently, but affect parameters within a die equally. 2) Within-die (WID) variation, which is also called intradie variation, across-chip variation, onchip variation, spatial variation, or spatial correlation,[1] describes the variation that affects process parameters at different locations of the same die differently. According to the scale of their causes, process variations can also be classified into the following two categories [14], [16]: 1) Systematic variation describes the deterministic portion of the variation. The physical cause of this variation is usually those identifiable deterministic features or patterns. For example, process conditions may vary randomly from wafer to wafer, but they may have a portion that is deterministically shared for all dies within the wafer. Similarly, process conditions may vary randomly from die to die, but a portion of them may be deterministically shared within a die. For example, interlayer dielectric-thickness variation is systematic and depends on layout density. 2) Random variation describes the variation that is independent of any other conditions. The physical cause of this variation is usually not well understood, and thus, it behaves more like a stochastic process. For example, discrete doping placement randomly changes MOSFET threshold voltage.

Both systematic and random variations need to be considered to accurately model the impact of process variations on designs. For example, the systematic-transistor channel-length variation can be more than 50% of its overall variations, and its root of causes includes through-pitch variation due to proximity (pitch) effects, through-process variation due to defocus condition, topography variation, mask variation, and etching [17]. Because systematic variations usually can be modeled accurately once a circuit's physical layout is known [9], [18], it can be corrected via techniques such as optical proximity correction, which postprocesses mask data so that the distortion of printed image caused by proximity environment of the designed shapes can be reduced [19]. In contrast, random variations are more like a stochastic process, and there is no clear trend or pattern to be predicted. It is the random process variation that permits us to treat designs statistically, including statistical timing analysis and optimization [1], [2].

We denote F as the measurable process parameter of interest, which can be either a physical parameter, like channel length, channel width, silicon oxide thickness, and wire thickness, or a parametric quantity, such as gate delay and threshold voltage.[2] Because of manufacturing-process variations, these process parameters are no longer fixed values. We model the parameter as a random variable, which is a complicated function of D2D systematic and random variations, and WID systematic and random variations. Conceptually, we can represent it as

$$F = h(Z_{\mathrm{D2D,\,sys}}, Z_{\mathrm{D2D,\,rnd}}, Z_{\mathrm{WID,\,sys}}, Z_{\mathrm{WID,\,rnd}}) \quad (1)$$

where $Z_{\mathrm{D2D,\,sys}}$ models the D2D systematic variation, $Z_{\mathrm{D2D,\,rnd}}$ models the D2D random variation, $Z_{\mathrm{WID,\,sys}}$ models the WID systematic variation, and $Z_{\mathrm{WID,\,rnd}}$ models the WID random variation. All variation components are further complicated functions of the manufacturing process, the feature's relative location in the wafer, the feature's relative location in the die, and the feature's local geometry patterns, to name just a few.

B. Process-Variation Decomposition

Assuming that the impact of each variation component is linear, we write (1) as

$$F = h_0 + h_1(Z_{\mathrm{D2D,\,sys}}) + h_2(Z_{\mathrm{D2D,\,rnd}}) + h_3(Z_{\mathrm{WID,\,sys}}) + h_4(Z_{\mathrm{WID,\,rnd}}) + X_r \quad (2)$$

where h_0 is a function that models the nominal value of F under nominal manufacturing conditions without any variation; h_1, h_2, h_3, and h_4 are the functions that model the impact of respective variation component (i.e., $Z_{\mathrm{D2D,\,sys}}$, $Z_{\mathrm{D2D,\,rnd}}$, $Z_{\mathrm{WID,\,sys}}$, and $Z_{\mathrm{WID,\,rnd}}$) on F, and X_r is a residual part that models the purely independent random variation that is not explainable by other variation components. The sum of h_1 and h_2 reflects the fluctuation of F caused by D2D variation, and the sum of h_3 and h_4 reflects the fluctuation of F caused by WID variation. The sum of h_1 and h_3 reflects the fluctuation of F caused by systematic variation, and the sum of h_2 and h_4 reflects the fluctuation of F caused by random variation

$$F_s = h_1(Z_{\mathrm{D2D,\,sys}}) + h_3(Z_{\mathrm{WID,\,sys}})$$
$$F_r = h_2(Z_{\mathrm{D2D,\,rnd}}) + h_4(Z_{\mathrm{WID,\,rnd}}) + X_r$$

[1] Spatial variation or spatial correlation is used more frequently in recent literature than the others.

[2] Without loss of generality, we use one generic-process parameter F in the following discussion. But, it is understood that the same techniques to be presented can be easily extended to multiple-process parameters.

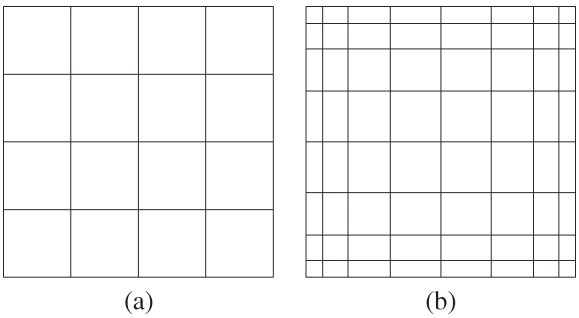

Fig. 1. Grid-based spatial-correlation models. (a) Uniform grids. (b) Nonuniform grids.

where F_s models the systematic variation of F, while F_r is a zero-mean random variable that models the random variation of F. Hence, we have

$$F = h_0 + F_s + F_r. \tag{3}$$

The variance of F, σ_F^2, is also called the overall chip variance.

III. Spatial-Correlation Modeling and Problem Formulations

It has been observed that devices that are physically close to each other are more likely to have similar characteristics than devices that are far apart. This phenomenon is captured by the modeling of spatial correlation. In the following, we introduce two ways to model the spatial correlation, each of which has its own value and applies to different process-variation scenarios.

A. Grid-Based Model of Spatial Correlation

1) Modeling: In this model, a set of grid cells is superimposed on top of the chip area, as shown in Fig. 1. It is assumed that the difference between process parameters only occurs for process parameters at different grid cells, and all process parameters within the same grid cell will have the same characteristics. In other words, the spatial correlation for process parameters within one grid cell is always one, and it is only interesting to know the spatial correlation between process parameters at different grid cells.

The grid-based spatial-correlation model can be adapted to handle more complicated variation scenarios by varying the number, size, and shape of grid cells. For example, it is believed that process control at chip center area is better than at chip boundaries; hence, spatial correlation at the chip center area is more uniform than at the boundaries. In this case, we can apply the gridding scheme, as shown in Fig. 1(b), where the center grid cells are coarser and less, while the boundary grid cells are finer and more. Hence, the grid-based model can easily capture the nonuniform spatial-correlation phenomena across the whole chip. Moreover, the shape of grids can be also nonrectangular, and for different process parameters, the gridding schemes may also be different.[3]

[3]The optimal way of gridding, including grid numbers, grid shapes, and grid sizes, can be decided under the guidance of good knowledge of manufacturing process, which is not addressed in this paper.

If we associate every grid cell i in the chip area with a random variable F_i and denote its variance as $\sigma_{F_i}^2$, then for the parameter of interests at two different grid cells i and j, the overall covariance between them is given by

$$\operatorname{cov}(F_i, F_j) \equiv \rho_{i,j} \cdot \sigma_{F_i} \cdot \sigma_{F_j} \tag{4}$$

where $\rho_{i,j}$ is the overall process correlation between process parameters at grid cell i and j.

For M number of chosen grid cells on the chip, we assume the joint-spatial variation $\mathbf{F} = (F_1, F_2, \ldots, F_M)^{\mathrm{T}}$ follows a multivariate Gaussian process with respect to their respective physical locations on the chip. To fully characterize the M-dimensional Gaussian distribution, we need to know the variance $\sigma_{F_i}^2$ for all grid cells and their corresponding correlation matrix $\mathbf{\Omega}$, as shown in (5)

$$\mathbf{\Omega} = \begin{bmatrix} 1 & \rho_{1,2} & \rho_{1,3} & \cdots & \rho_{1,M} \\ \rho_{1,2} & 1 & \rho_{2,3} & \cdots & \rho_{2,M} \\ \rho_{1,3} & \rho_{2,3} & 1 & \cdots & \rho_{3,M} \\ \cdot & & & & \cdot \\ \rho_{1,M} & \rho_{2,M} & \rho_{3,M} & \cdots & 1 \end{bmatrix}. \tag{5}$$

A valid correlation matrix must be positive semidefinite by its definition [20].

2) Problem Formulation: Based on the grid-spatial-correlation model, we propose the following problem formulation.

Formulation 1: Extraction of Spatial-Correlation Matrix: Given M number of grid cells on a chip to model the spatial correlation, and noisy measurement data for the parameter of interest at these grid cells, extract the overall process variation at every grid cell $\sigma_{F_i}^2$ and their corresponding spatial-correlation matrix Ω, as shown in (5), so that it not only accurately captures the underlying process-variation model, but the extracted correlation matrix Ω is always positive semidefinite.

Extracting a valid correlation matrix is of practical significance. For example, SSTA tools such as [1], which are based upon principle-component analysis, require that the spatial-correlation matrix must be valid and known *a priori*.

Even though the grid-based spatial-correlation model is intuitively simple and easy to use, it has its own limitations. The foremost one is the inherent accuracy-versus-efficiency issue because of its fundamental assumption, which states that all parameters of interests within one grid cell have the same characteristics. To justify such an assumption, the size of each grid cell cannot be too large, which in turn increases the total number of grid cells required for modeling. From extraction point of view, on the one hand, the more number of grid cells, the more number of measurement sites within each chip, hence, the more expensive to extract such a model. On the other hand, the physical limitation of measurement devices also prevents the grid-cell size from being too small; otherwise, the measurement probe would not fit into one grid cell. Because of these inherent limitations of the grid-based modeling approach, in the next section, we propose a more flexible approach to model spatial correlation.

B. Gridless-Based Model of Spatial Correlation

1) Modeling: Because the systematic variation is more like a deterministic variation [14], we lump it with the nominal value h_0, i.e.,

$$f_0 = h_0 + F_s \tag{6}$$

where f_0 is the mean value of F with the systematic variation considered. The extraction of the mean value f_0 is relatively easy and is essentially done through averaging. For example, [9] has presented a methodology to extract the intradie systematic variation of critical dimension (CD or effective channel length) by averaging out measurement of CD at different locations of the same die.

In the following, we mainly concern ourselves in extracting the random variation parts F_r. This is a more challenging task, because simply taking averaging of measurements would not give us any useful information on the zero-mean random-process variations' characteristics. Towards this end, we rewrite the random variation F_r as

$$F_r = X_g + X_s + X_r$$

where X_g models the interchip global variation that affects all features within the same chip equally but is different among different chips obtained from different lots, wafers, or even the same wafer. In other words, $X_g = h_2(Z_{\mathrm{D2D,\,rnd}})$. Intrachip spatial correlation is modeled by X_s, which is different for features at different locations within the same chip. In other words, $X_s = h_4(Z_{\mathrm{WID,\,rnd}})$. Therefore, we have

$$F = f_0 + F_r = f_0 + X_g + X_s + X_r. \tag{7}$$

The three types of random variations, X_g, X_s, and X_r, are independent by definition. Hence, the variance of F_r is given by

$$\sigma_{F_r}^2 = \sigma_G^2 + \sigma_S^2 + \sigma_R^2 \tag{8}$$

where σ_G^2, σ_S^2, and σ_R^2 are the variances of X_g, X_s, and X_r, respectively. When the systematic variation is excluded, the overall chip variance is equivalent to the random variance, i.e., $\sigma_F^2 = \sigma_{F_r}^2$.

We model the random part of process variation F_r as a homogeneous and isotropic random field, whose formal definition is introduced as follows.

Definition 1: **Random Field** is a real random function $F(x, y)$ of position (x, y) in the two-dimensional space $\mathcal{R}^2$.

Definition 2: **Homogeneous** and **Isotropic Random Field** is a random field $F(x, y)$ whose mean and variance are constants and whose correlation function ρ (x_i, x_j, y_i, and y_j) between any two points depends only on the distance v between them, i.e.,

$$\rho(x_i, x_j, y_i, y_j) = \rho(v_{i,j}) \tag{9}$$

where $v_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$.

If the spatial variation follows a homogeneous and isotropic random field, then the same distance $v_{i,j}$ always corresponds to the same $\rho(v_{i,j})$, regardless of their locations. Therefore, for simplicity, we denote $\rho(v_{i,j})$ as $\rho(v)$ in the following whenever there is no ambiguity.

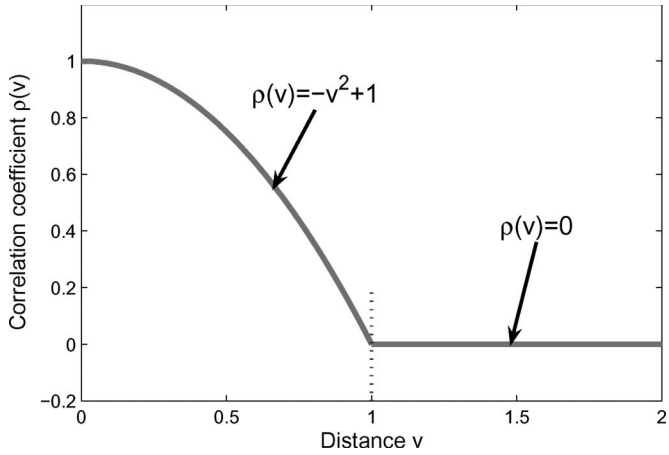

Fig. 2. Monotonically decreasing function that is not a valid spatial-correlation function.

Note that the overall process variation F, as shown in (3), that includes both systematic variation and random variation, does not necessarily follow a homogeneous and isotropic random field. But, if we only look at the random-variation part, then the physical properties of the random-variation part F_r would be very likely to follow a homogeneous and isotropic random field, particularly, when the manufacturing process becomes mature and stable.

2) Valid Spatial-Correlation Function: Formally, a valid spatial-correlation function $\rho(v)$ is a function such that the correlation matrix generated from $\rho(v)$ for arbitrary number of points on the two-dimensional space is always positive semidefinite.

In its simplest way, a valid spatial-correlation function should satisfy the following necessary but not sufficient conditions[4]:

$$\rho(0) = 1 \tag{10}$$

$$0 \le \rho(v) \le 1 \tag{11}$$

$$\rho'(v) \le 0. \tag{12}$$

Equations (10) and (11) are required by the definition of correlation coefficient [20]. The interpretation of (12) is that the spatial correlation is a monotonically decreasing function of distance, i.e., as devices become further apart, the correlation between them becomes smaller. The correlation distance $\overline{v}$ is the distance beyond which the spatial correlation $\rho(v)$ becomes sufficient small and can be approximated as zero, i.e., $\rho(v) \approx 0$ for all $v \ge \overline{v}$. For simplicity, in the following, when we describe the correlation function, we only give the function form for any $v \in [0, \overline{v}]$ whenever there is no ambiguity. For example, Fig. 2 shows a piecewise monotonically decreasing function, and the function form is $\rho(v) = -v^2 + 1$ for any $v \in [0, \overline{v}]$ with $\overline{v} = 1$.

Contrary to the common wisdom, we show that not all monotonically decreasing functions qualify for the spatial-correlation function. For example, for the function as shown in

[4]In the context of process variation, we are only interested in the spatial correlation that is nonnegative.

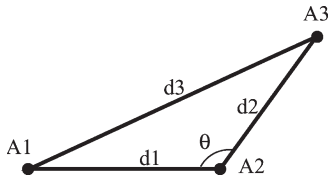

Fig. 3. Any three points on the die.

Fig. 2, if we assume $\rho(v)$ is a valid spatial-correlation function, then the correlation matrix $\mathbf{\Omega}$ between any three points on the die as shown in Fig. 3 can be built as

$$\Omega = \begin{bmatrix} 1 & \rho(d_1) & \rho(d_3) \\ \rho(d_1) & 1 & \rho(d_2) \\ \rho(d_3) & \rho(d_2) & 1 \end{bmatrix}. \tag{13}$$

If $d_1 = 31/32$, $d_2 = 1/2$, and $d_3 = 1/2$, under $\rho(v) = -v^2 + 1$, we obtain a correlation matrix[5]

$$\Omega = \begin{bmatrix} 1 & 0.0615 & 0.75 \\ 0.0615 & 1 & 0.75 \\ 0.75 & 0.75 & 1 \end{bmatrix}. \tag{14}$$

However, it is easy to show that the resulting matrix is not a valid correlation matrix, as the smallest eigenvalue of this matrix is -0.0303, implying that this matrix is not positive semidefinite.

Recall that a matrix is a positive semidefinite matrix if and only if its every principal submatrix has a nonnegative determinant, whereas the principal submatrices are formed by removing row–column pairs from the original symmetric matrix [21]. Therefore, in order for Ω, in (13), to be a positive semidefinite matrix, we require the following two equations to hold:

$$1 - \rho(d_1)^2 \geq 0 \tag{15}$$

$$1 + 2\rho(d_1)\rho(d_2)\rho(d_3) \geq \rho(d_1)^2 + \rho(d_2)^2 + \rho(d_3)^2. \tag{16}$$

Equation (15) is automatically satisfied following (11). Therefore, we only need to check (16). Plugging the values of $d_1 = 31/32$, $d_2 = 1/2$, and $d_3 = 1/2$ with $\rho(v) = -v^2 + 1$ into (16), we can see that it violates the constraints of (16). This explains why function $\rho(v) = -v^2 + 1$ is not a valid spatial-correlation function.

This leads us to the question of "what type of monotonic decreasing functions qualify to be a valid spatial-correlation function?" To answer this question, we introduce the following theorem.

Theorem 1: A necessary and sufficient condition for the function $\rho(v)$ to be a valid spatial-correlation function of a homogeneous and isotropic random field is that it can be represented in the form of

$$\rho(v) = \int_0^\infty J_0(\omega v) d\left(\Phi(\omega)\right) \tag{17}$$

[5]The three points form a triangle, which imposes constraints on the possible choices of d_1, d_2, and d_3, i.e., $d_1 + d_2 > d_3$, $d_2 + d_3 > d_1$, and $d_1 + d_3 > d_2$.

where $J_0(t)$ is the Bessel function of order zero and $\Phi(\omega)$ is a real nondecreasing function on $[0, \infty)$ such that for some nonnegative p

$$\int_0^\infty \frac{d\Phi(\omega)}{(1+\omega^2)^p} < \infty. \tag{18}$$

Proof: See [22] for the proof. ■

Based on the above theorem, we derive the following two corollaries.

Corollary 1: The monotonically decreasing exponential function (19) and double exponential function (20), i.e.,

$$\rho(v) = \exp(-bv) \tag{19}$$

$$\rho(v) = \exp(-b^2 v^2) \tag{20}$$

are valid spatial-correlation functions. The constant b is a parameter that regulates the decaying rate of the correlation function with respect to distance v. The correlation distance for the two functions are infinity, i.e., $\overline{v} = \infty$.

Proof: By way of construction, we find that $\Phi(\omega) = 1 - (1/\sqrt{1 + \omega^2/b^2})$ and $\Phi(\omega) = 1 - \exp(-\omega^2/4b^2)$ satisfy the conditions as specified in Theorem 1. Plugging them into (17), we obtain the corresponding correlation functions as (19) and (20), respectively. In other words, the exponential function (19) and double exponential function (20) are valid spatial-correlation functions [23]. ■

Corollary 2: The monotonically decreasing linear function in the form of

$$\rho(v) = -av + b, \qquad \forall v \in [0, \overline{v}] \tag{21}$$

with $\overline{v} \leq b/a$ is not a valid spatial-correlation function, where a and b are two positive numbers.

Proof: To prove that (21) is not a valid spatial-correlation function, all we need to do is to find a counter example, by which a correlation matrix generated from it is not positive semidefinite. One such counter example was given in [24], which convincingly shows that the monotonically decreasing linear function as shown in (21) is not a valid spatial-correlation function. ■

The implication of Corollary 2 is interesting to note, because intuitively people may think that the monotonically decreasing linear function is valid for spatial-correlation modeling, and the work, as shown in [10], did apply it to real wafer data. But, Corollary 2 tells us that such a practice is not correct.

In general, it is difficult to check whether an arbitrary function form is a valid correlation function [24]. For example, for an arbitrary piecewise linear function, i.e, with arbitrary number of linear segments and each with arbitrary slopes, we cannot provide (and fail to find) any theoretical proof showing whether or not it is valid. But, we at least can say for sure that not all piecewise linear function is valid, because, as shown in [24], for some particular piecewise linear function, the spatial-correlation matrix generated from it is not valid. In the work of [11], the authors proposed to use a piecewise linear function to model the spatial-correlation function. However, there is no guarantee that the so-obtained piecewise linear function is

valid. The authors of [11] also did not provide any theoretical justification for their approach.

3) Problem Formulation: When the spatial variation follows a homogeneous and isotropic random field, we propose the following second problem formulation.

Formulation 2: Extraction of Spatial-Correlation Function: Given noise-measurement data for the parameter of interest with possible inconsistency, extract the interchip global-variation component σ_G^2, the intrachip spatial-variation component σ_S^2, the random-variation component σ_R^2, and the spatial-correlation function $\rho(v)$, so that the extracted variation components accurately capture the underlying variation model, and the spatial-correlation function is always a valid correlation function satisfying condition (17).

If the spatial variation is modeled as a homogeneous and isotropic random field in a two-dimensional space $\mathcal{R}^2$, then for the parameter of interest at arbitrary two different points, their covariance is

$$\operatorname{cov}(F_i, F_j) = \operatorname{cov}(X_g, X_g) + \operatorname{cov}(X_{s,i}, X_{s,j}) \quad (22)$$

$$= \sigma_G^2 + \rho(v)\sigma_{vS}^2 \quad (23)$$

where $\rho(v)$ is the spatial-correlation coefficient between two locations that are v distance apart. In other words, we can characterize the process variation by extracting the interchip global variation σ_G^2, intrachip spatial variation σ_S^2, and the correlation function $\rho(v)$.

For the parameter of interest at two different locations with distance of v, the overall process correlation between them is thus given by

$$\rho_v \equiv \frac{\operatorname{cov}(F_i, F_j)}{\sigma_{F_i}\sigma_{F_j}} \quad (24)$$

$$= \frac{\sigma_G^2 + \rho(v)\sigma_S^2}{\sigma_G^2 + \sigma_S^2 + \sigma_R^2}. \quad (25)$$

Because the spatial correlation $\rho(v)$ is a function of the distance v so is the overall process correlation ρ_v. As $\rho(v)$ is homogeneous and isotropic so is ρ_v. Because of the one-to-one correspondence between spatial correlation $\rho(v)$ and the overall process correlation ρ_v, extracting the spatial-correlation function $\rho(v)$ is equivalent to extracting the overall process correlation function ρ_v.

In Fig. 4, we show a possible curve for the overall correlation ρ_v as a function of the distance v, as given by (25). According to Fig. 4, the total correlation can be divided into three parts: Part G is the correlation caused by the interchip global variation; part S is the correlation caused by the intrachip spatial correlation; and part R is caused by the purely uncorrelated random variation. We can see that the overall process correlation ρ_v starts to settle at a constant value when the distance becomes large enough (greater than the correlation distance $\overline{v}$), which means that even for devices from the same chip that are far apart, there is still some correlation between them due to their shared global variations. We can also see that there is a sudden drop from one for ρ_v at distance zero. The cause for that drop is the purely uncorrelated random variation, such that even for devices that are very close to each other, they are still not perfectly correlated. Perfect correlation $(\rho_v = 1)$ only occurs when the two devices are in fact the same device.[6]

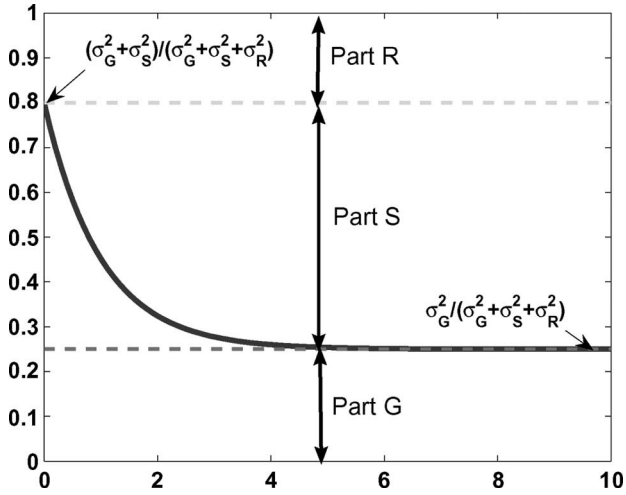

Fig. 4. Possible curve for the overall process correlation according to (25) in the absence of measurement noise (theoretically ideal case).

C. Extraction Setup

To experimentally characterize the process variation, we obtain N samples of a chip, and choose M number of sites on each chip, where measurement is conducted. The sites are denoted as (x_i, y_i), and the distance between any two sites is denoted as $v_{i,j}$. We denote each measurement of the parameter of interest F as $f_{k,i}$ for the kth chip on the ith site.

Note that in order to obtain these measurement data, it usually requires careful design of test structures, placement of test structures spanning a range of areas on the die, and measurement procedures to collect data. Those details are beyond the scope of this paper, and interested readers are referred to [16] for more information.

In the following, we present techniques to solve the above two problem formulations as discussed in Sections III-A and III-B, respectively. We first solve the extraction of spatial-correlation function in Section IV, then solve the extraction of spatial-correlation matrix in Section V.

IV. Extraction of Valid Spatial-Correlation Function

A. Global Variation Extraction

We treat each measurement of the parameter of interest F as a sampling of the quantity in (7). Given N samples of a chip and M number of measurement sites on each chip, we group the measured data $f_{k,i}$ by their chip locations as follows: $f_{k,\cdot} = [f_{k,1}, \ldots, f_{k,M}]$ for $k = 1$ to N, or by their site locations as follows: $f_{\cdot,i} = [f_{1,i}, \ldots, f_{N,i}]$ for $i = 1$ to M. For better presentation, we denote the actual variance as σ^2

[6]Note that a similar plot showing the trend of the overall process correlation with respect to distances has also been empirically observed in [11] based on wafer-scale measurements. But, the authors of [11] did not provide a theoretical explanation of this phenomena as we do in this paper.

with an upper case letter in subscript, like σ_G^2 for the global-variation component, and denote the extracted variance as σ^2 with a lower case letter in the subscript, like σ_g^2 for the extracted global-variation component.

We approximate the overall chip variance σ_F^2 by computing the unbiased sample variance [20] of $f_{k,i}$ as follows[7]:

$$\sigma_F^2 \approx \sigma_f^2 = \frac{1}{M(N-1)} \sum_i \left(\sum_k f_{k,i}^2 - \frac{\left(\sum_k f_{k,i}\right)^2}{N} \right). \tag{26}$$

For all samples of the parameter of interest F within a particular chip c, because the interchip global variation X_g changes the value of parameter for all samples with the same chip by the same amount, the overall within-chip variance is, thus, given by

$$\sigma_{F_c}^2 = \sigma_S^2 + \sigma_R^2. \tag{27}$$

We estimate the overall within-chip variation by computing the unbiased sample variance [20] of $f_{k,\cdot}$ as follows[8]:

$$\sigma_{F_c}^2 \approx \sigma_{f_k}^2 = \frac{1}{M-1} \left(\sum_i f_{k,i}^2 - \frac{\left(\sum_i f_{k,i}\right)^2}{M} \right). \tag{28}$$

For different $f_{k,\cdot}$, we may get different estimation of $\sigma_{F_c}^2$ caused by inconsistent measurement. To improve the accuracy, we estimate the overall within-chip variance by taking the average value of $\sigma_{f_k}^2$. We denote the resulting average value as $\sigma_{f_c}^2 \approx \sigma_{F_c}^2$.

Knowing the estimation of the overall chip variance σ_f^2 and the overall within-chip variance $\sigma_{f_c}^2$, we extract the interchip global variation by

$$\sigma_G^2 = \sigma_F^2 - \sigma_{F_c}^2 \approx \sigma_g^2 = \sigma_f^2 - \sigma_{f_c}^2. \tag{29}$$

B. Spatial-Correlation Extraction

For any two different sets of $f_{\cdot,i}$ and $f_{\cdot,j}$ at two different sites that are v distance apart, we estimate the covariance of F_i and F_j by computing the unbiased sample covariance [20] of $f_{\cdot,i}$ and $f_{\cdot,j}$ as follows:

$$\operatorname{cov}(F_i, F_j) \approx \operatorname{cov}(f_{\cdot,i}, f_{\cdot,j}) \tag{30}$$

$$= \frac{\sum_k f_{k,i} f_{k,j}}{N-1} - \frac{\sum_k f_{k,i} \sum_k f_{k,j}}{N(N-1)}. \tag{31}$$

For simplicity, we also denote $\operatorname{cov}(f_{\cdot,i}, f_{\cdot,j})$ as $\operatorname{cov}(v)$ to show that it is a function of two points that are v distance apart.

[7]The advantage of using the unbiased sample variance over sample variance is that it will not over or under-estimate the true quantity. In practice, the true or exact variance of a population is not known *a priori* and has to be computed based on samples. Unbiased sample variance is good at estimating the true variance in this case. In contrast, the sample variance merely measures the variance for the given finite number of samples, hence, it is a biased estimator of the true variance. For more information about unbiased sample variance and variance, please refer to [20].

[8]Note, to use (28) for unbiased sample variance, we need to choose measurement data from a subset of sites i whose spatial correlation is zero. Otherwise, without using (28) and (29), we need to treat σ_g^2 as an unown in solving (33).

According to (23) and (29), we estimate the product of spatial variation σ_S^2 and spatial correlation $\rho(v)$ as follows:

$$\sigma_S^2 \cdot \rho(v) = \operatorname{cov}(F_i, F_j) - \sigma_G^2 \approx \operatorname{cov}(v) - \sigma_g^2. \tag{32}$$

Because $\rho(v)$ is a function of v, we need to compute $\rho(v)$ for different pairs of sites with different distances in order to obtain the full description of $\rho(v)$. But, there are two challenges in doing that: 1) We do not know the exact value of spatial variation σ_S^2. 2) Because of unavoidable measurement errors, the data set computed as above may not be consistent. Therefore, in the following, we propose a robust technique to find the spatial-correlation function $\rho(v)$ and σ_S^2 accurately. Moreover, the resulting $\rho(v)$ is guaranteed to be a valid spatial-correlation function.

Given the data set $[v, \operatorname{cov}(v)]$ as computed from (31), we formulate the robust spatial-variation extraction problem as the following optimization problem:

$$\begin{aligned} \min_{\Phi, \sigma_s^2} : \quad & \left\| \sigma_s^2 \int_0^\infty J_0(\omega v) d\left(\Phi(\omega)\right) - \operatorname{cov}(v) + \sigma_g^2 \right\| \\ s.t. \quad & \sigma_s^2 \le \sigma_{f_c}^2, \\ & \int_0^\infty \frac{d\Phi(\omega)}{(1+\omega^2)^p} < \infty. \end{aligned} \tag{33}$$

In other words, we find a valid spatial-correlation function by solving a constrained nonlinear optimization problem, so that the resulting spatial-correlation function minimizes the total error with respect to measurement data. After obtaining $\Phi(\omega)$, we plug it into (17) to obtain the valid spatial-correlation function $\rho(v)$.

The previous problem formulation is very general and applies to any real nondecreasing function $\Phi(\omega)$. For practical use, however, there is no need to enumerate all possible choices of $\Phi(\omega)$ in order to find the optimal $\rho(v)$. Moreover, as we have discussed in Section III-B2, it is also difficult to check the validity of an arbitrary spatial-correlation function.

Therefore, to make the problem tractable, we can approximate the experimentally measured correlation function with a function selected from a family of functions that are proved to be valid spatial-correlation functions. To serve such a purpose, it is sufficient to chose a family of functions $\Phi(\omega)$ so that the $\rho(v)$ obtained from (17) contains a rich set of functions for the purpose of modeling spatial correlation.

It has been shown in [23] that by choosing a proper family function of $\Phi(\omega)$, we obtain a very general family of spatial-correlation functions

$$\rho(v) = 2 \left(\frac{bv}{2} \right)^{s-1} K_{s-1}(bv) \Gamma(s-1)^{-1} \tag{34}$$

where K is the modified Bessel function of the second kind, Γ is the gamma function, and b and s are two real parameter numbers that regulate the shape of the function. By varying b and s, we obtain different spatial-correlation functions.

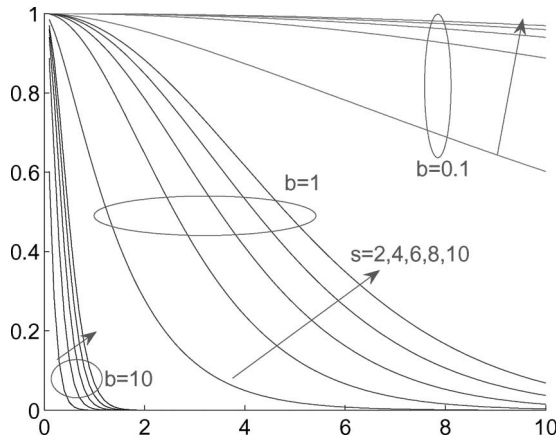

Fig. 5. Correlation functions generated from (34).

For example, the exponential function, as shown in (17), can be generated from (34) by choosing $s = 3/2$.

To show that the function of (34) indeed provides us a rich set of correlation functions that suffice for our spatial-correlation modeling, we plot the function of (34) under different parameters of b and s. Fig. 5 shows a few samples of correlation functions generated from (34) by setting b to be 0.1, 1, and 10, and varying s from 2 to 10 with a step size of two. From the figure, it is shown that the correlation function (34) indeed can generate a rich gamut of correlation functions for the purpose of spatial-correlation modeling.

Without loss of generality, in the following, (34) will be used as the candidate[9] correlation function in (33). Moreover, two-norm is used as a measure of the objective function in (33). Therefore, we rewrite the optimization problem as given in (33) as follows:

$$\min_{b,s,\sigma_s^2}: \sum \left[2\sigma_s^2 \left(\frac{bv}{2}\right)^{s-1} K_{s-1}(bv)\Gamma(s-1)^{-1} - \mathrm{cov}(v) + \sigma_g^2 \right]^2$$
$$s.t. \qquad \sigma_s^2 \le \sigma_{f_c}^2. \tag{35}$$

This is a constrained nonlinear least square problem, and we can solve it efficiently via any nonlinear least square technique [25]. Note that problem (35) is not a convex problem in general; hence, we cannot guarantee to find a global optimal solution. But as this kind of least square minimization problem is well-studied in the literature, good solvers are available to find a solution with high quality. Our experimental results to be presented also confirms this argument. Moreover, as all nonlinear minimization engines are sensitive to the initial guess, obtaining a high-quality solution sometimes may require us to try different initial guesses.

[9]Function (34) is chosen over the exponential (19) or double exponential (20) function as a candidate spatial-correlation function in this paper. The reason is that it has more parameters (b and s) and contains the exponential function as a special case (with $s = 3/2$). This gives us considerably more flexibility to fit the data but still with reasonable complexity. Apparently, other choices of candidate functions are possible. But, we have to be careful in assuring that the candidate functions are valid spatial correlation function. As pointed out by [24], it is always simpler and safer to use those "approved" valid spatial-correlation functions, as testing the validity of an arbitrary function form (such as linear and piecewise linear) is almost always time consuming and difficult.

1	Extract global variation σ_g^2 by (29);
2	Solve (35) to obtain σ_s^2 and b and s;
3	Extract $\overline{\rho(v)}$ by plugging b and s into (34);
4	Extract random variation σ_r^2 by (36);
5	Extract overall process correlation by (25);

Fig. 6. Algorithm for characterization of process variation.

After solving the above problem, we obtain the estimated spatial-variation component $\sigma_S^2 \approx \sigma_s^2$ and the parameter b and s. By plugging b and s into (34), we obtain the estimated spatial-correlation function $\overline{\rho(v)} \approx \rho(v)$. Therefore, we have obtained all information about the spatial-variation component: both the variance of spatial variation and the spatial-correlation function.

C. Overall Algorithm

The overall algorithm for characterizing the process variation is summarized, as shown in Fig. 6.

We first extract the global-variation component σ_g^2 by using (29). We then solve the nonlinear least square optimization problem as defined in (35) to obtain the spatial-variation component σ_s^2 and the parameter of b and s that define the spatial-correlation function for a homogeneous and isotropic random field, as shown in (34). According to (8), we extract the random-variation component by using the following formula:

$$\sigma_R^2 = \sigma_F^2 - \sigma_G^2 - \sigma_S^2 \approx \sigma_r^2 = \sigma_f^2 - \sigma_g^2 - \sigma_s^2. \tag{36}$$

By plugging all variation components into (25), we obtain the overall process correlation at any distance.

V. Extraction of Spatial-Correlation Matrix

A. Overall Algorithm

We are given measurement data with possible inconsistency (e.g., missing measurement data and noisy measurement) for a number of points on each chip and some samples of the same chip. We extract the overall process spatial correlation as follows.

We first estimate the covariance between any two points that have measurements on the same set of N chips by (31). We then estimate the variance of each point with N measurements by computing its unbiased sample variance [20] as follows:

$$\sigma_{F_i}^2 \approx \sigma_{f_i}^2 = \frac{1}{N-1}\left(\sum_k f_{k,i}^2 - \frac{\left(\sum_k f_{k,i}\right)^2}{N}\right). \tag{37}$$

By plugging the estimated $\sigma_{f_i}^2$ and $\sigma_{f_j}^2$ and $\mathrm{cov}(f_{.,i}, f_{.,j})$ from (31) into (24), we obtain the estimated overall process correlation coefficient

$$\rho_{i,j} = \frac{\mathrm{cov}(F_i, F_j)}{\sigma_{F_i}\sigma_{F_j}} \approx \frac{\mathrm{cov}(f_{.,i}, f_{.,j})}{\sigma_{f_i}\sigma_{f_j}}. \tag{38}$$

For the given M points of interest, we have $M(M-1)/2$ number of pairs of points F_i and F_j and the corresponding $M(M-1)/2$ number of estimated correlation coefficients $\rho_{i,j}$.

1 Compute $cov(f_i, f_j)$ by (31);
2 Compute $\sigma^2_{f_i}$ by (37);
3 Compute $\rho_{i,j}$ by (38);
4 Compute A by assembling $\rho_{i,j}$ into (5);
5 Compute Ω via the modified alternative projection algorithm;

Fig. 7. Algorithm for extracting a valid spatial-correlation matrix.

Putting all $\rho_{i,j}$ into (5), we obtain the estimated overall process spatial-correlation matrix $A \approx \Omega$.

Note that in order for the above estimated A to qualify for a correlation matrix, it has to be a positive semidefinite matrix. But, we cannot guarantee that such a property would hold automatically for the resulting A due to the unreliable (or inconsistent) measurement data. We solve this problem by employing the modified alternative-projection algorithm to be presented in the next section to robustly extract a valid correlation matrix Ω from the unreliable measurement data.

The overall algorithm for extracting a valid spatial-correlation matrix is summarized as follows in Fig. 7.

B. Modified Alternative-Projection Algorithm

The robust extraction of a consistent correlation matrix problem can be formulated as the following optimization problem. For a given symmetrical matrix A with elements $a_{i,j}$ between zero and one, find a correlation matrix Ω that is mostly close to A. Mathematically, the closeness can be measured via the distance between two matrices, i.e.,

$$\min_{\Omega}: \quad \|A - \Omega\| \tag{39}$$

$$s.t.: \quad \Omega \in \text{correlation matrix}. \tag{40}$$

We use the weighted Frobenius norm to measure the distance between two matrix. Recall that the Frobenius norm is defined as $\|A\|_F^2 = \sum a_{i,j}^2$. One of the weighted Frobenius norms is the W-norm as defined by

$$\|A\|_W = \|W^{1/2} A W^{1/2}\|_F \tag{41}$$

where W is a symmetric positive definite matrix.

This problem is also called the nearest correlation matrix problem [26], or the least squares covariance-adjustment problem [27]. As a proof of concept, we solve this problem by employing the modified alternative-projection algorithm proposed in [26] because of its ease of implementation. The idea is to iteratively project the symmetric matrix A onto two convex sets alternatively, and at the end of the iteration, the final projected matrix is the solution to the optimization problem as defined in (39).

We first define the sets

$$U = \{Y = Y^{\mathrm{T}} \in \mathbf{R}^{n\times n} : y_{ii} = 1\} \tag{42}$$

$$S = \{Y = Y^{\mathrm{T}} \in \mathbf{R}^{n\times n} : Y \geq 0\} \tag{43}$$

where the notation $Y \geq 0$ means that Y is positive semidefinite. Our desired correlation matrix Ω, as shown in (39), is a matrix

$\Delta S_0 = 0$, $Y_0 = A$
for k=1,2,...
  $R_k = Y_{k-1} - \Delta S_{k-1}$
  $X_k = P_S(R_k)$
  $\Delta S_k = X_k - R_k$
  $Y_k = P_U(X_k)$
end
$\Omega = Y_k$

Fig. 8. Modified alternative-projection algorithm.

that is in the intersection of U and S and has the shortest distance to A in a weighted Frobenius norm. Since S and U are both closed convex sets, so is their intersection. It thus follows from standard results in approximation theory that the minimum Ω in (39) is obtainable and unique.

Moreover, for a symmetric matrix $A \in \mathbf{R}^{n\times n}$ with spectral decomposition (or eigenvalue decomposition) $A = QDQ^{\mathrm{T}}$, where $D = \mathrm{diag}(\lambda_i)$ and Q is orthogonal, we introduce the following notations:

$$A_+ = Q\mathrm{diag}\left(\max(\lambda_i, 0)\right) Q^{\mathrm{T}}. \tag{44}$$

We denote $P_U(A)$ and $P_S(A)$ as the projections of A onto U and S, respectively. Then, for a given W-norm, $P_U(A)$ can be computed analytically via

$$P_U(A) = A - W^{-1}\mathrm{diag}(\theta_i)W^{-1} \tag{45}$$

where $\theta = [\theta_1, \ldots, \theta_n]^{\mathrm{T}}$ is the solution of the linear system

$$(W^{-1} \circ W^{-1})\theta = \mathrm{diag}(A - I) \tag{46}$$

where $\circ$ denotes the Hadamard product: $A \circ B = (a_{i,j}b_{i,j})$, i.e., elementwise matrix multiplication.

For a given W-norm, $P_S(A)$ can also be computed analytically via

$$P_S(A) = W^{-1/2}\left((W^{1/2}AW^{1/2})_+\right)W^{-1}. \tag{47}$$

When the W-norm is taken as the identity I, i.e., the unweighted Frobenius norm, $P_U(A)$ is simply as

$$P_U(A) = (p_{ij}) \tag{48}$$

with $p_{ij} = a_{ij}$ for all $i \neq j$ and $p_{ij} = 1$ for all $i = j$. For $P_S(A)$, it is simply as

$$P_S(A) = A_+ = Q\mathrm{diag}\left(\max(\lambda_i, 0)\right) Q^{\mathrm{T}}. \tag{49}$$

The following modified alternative-projection algorithm, as shown in Fig. 8, can be used to solve the nearest correlation-matrix problem, as defined in (39).

It has been proven that when $k \to \infty$, both X_k and Y_k converge to the desired correlation matrix Ω. Moreover, it has been theoretically shown that the convergence of the alternative-projection algorithm is linear [26]. This conclusion has also been experimentally verified in Section VI-B.

Among many possible choices, the following convergence condition can be used in Fig. 8 to stop the loop

$$\max\left\{\frac{\|X_k - X_{k-1}\|}{\|X_k\|}, \frac{\|Y_k - Y_{k-1}\|}{\|Y_k\|}, \frac{\|Y_k - X_k\|}{\|Y_k\|}\right\} \leq \epsilon$$

where ϵ is a small tolerance number (say $\epsilon = 10^{-8}$).

VI. Experiment Results

We employ a Monte Carlo model of measurement to verify the robustness and accuracy of our extraction algorithms in this paper. One of the advantages of using Monte Carlo simulation is that it allows us to simulate different variation scenarios and measurement settings that are difficult to control in reality. By comparing the extracted variation components with the known variation components used in the Monte Carlo model, we can quantitatively examine how robust and how accurate our extraction algorithms are in the presence of different amount of measurement errors. Such a study is useful, because it provides us the confidence in applying the algorithms to real wafer measurement.

A. Extraction of Valid Spatial-Correlation Function

In this paper, the Monte Carlo model is based on a valid correlation function $\rho(v)$ that follows a homogeneous and isotropic random field, but with different variation amounts for the three variation components (σ_G^2, σ_S^2, and σ_R^2). We simulate the measurement process by generating a set of measurement data from N number of sample chips and M number of measurement sites on each chip. To model the reality due to measurement error, we add a Gaussian noise with different variation amounts during the Monte Carlo sampling. By applying the algorithm, as shown in Fig. 6, we extract the global-variation component σ_g^2, random-variation component σ_r^2, spatial-variation component σ_s^2, and parameter of b and s that define the spatial-correlation function $\overline{\rho(v)}$ for a homogeneous and isotropic random field, as shown in (34). By plugging all variation components into (25), we obtain the overall process correlation at any distance. We measure the accuracy of our extraction algorithm for the global variation and spatial variation, but not the random variation as it is indistinguishable from the added measurement noise. For the global-variation component, the relative error is given by

$$\text{err}\left(\sigma_G^2\right) = \frac{\sigma_g^2 - \sigma_G^2}{\sigma_G^2}. \tag{50}$$

For the spatial-variation component, the relative error is given by

$$\text{err}\left(\sigma_S^2\right) = \frac{\sigma_s^2 - \sigma_S^2}{\sigma_S^2}. \tag{51}$$

Moreover, for the spatial-correlation function, the relative error is given by

$$\text{err}\left(\rho(v)\right) = \frac{\left\|\overline{\rho(v)} - \rho(v)\right\|}{\|\rho(v)\|}. \tag{52}$$

TABLE I
Process-Variation Extraction

N	M	$Noise$	$err(\sigma_G^2)$	$err(\sigma_S^2)$	$err(\rho(v))$
2000	60	10%	0.4%	-1.9%	2.0%
		50%	0.3%	-2.8%	2.7%
		100%	0.3%	-2.6%	3.7%
1500	60	10%	4.1%	2.5%	0.9%
		50%	3.9%	2.1%	1.0%
		100%	3.8%	2.0%	1.2%
1000	60	10%	7.5%	1.2%	1.0%
		50%	7.2%	1.0%	1.0%
		100%	6.9%	1.4%	1.0%
500	60	10%	17.8%	10.9%	6.6%
		50%	18.3%	6.1%	4.8%
		100%	18.6%	4.7%	3.1%
1000	50	10%	6.5%	0.8%	2.8%
		50%	5.7%	-0.4%	3.0%
		100%	5.1%	-3.0%	3.5%
	40	10%	8.6%	-4.1%	6.5%
		50%	8.7%	-3.9%	7.0%
		100%	8.9%	-2.3%	8.4%

From statistical theories, we know that if we have more measurement data, we have more confidence in the accuracy of statistics obtained from measurements. In reality, however, measurement of chips is usually very time-consuming and expensive. Therefore, it is desirable to attain similar accuracy yet with as few number of measurement data as possible. A robust extraction algorithm helps to achieve that goal.

We report experimental results in Table I, where N is the number of sample chips, M is the number of measurement sites, noise is the amount of random noise added into the Monte Carlo model in terms of the total variation $(\sigma_G^2 + \sigma_S^2 + \sigma_R^2)$. The product of N and M gives the total number of measurements.

According to Table I, we see that our algorithm is very accurate in extracting different variation components, yet very robust to different amount of random noise. For example, with $N = 2000$, $M = 60$, and Noise $= 10\%$, our extracted results have about 0.4% error for the global variation, 1.9% error for the spatial variation, and 2.0% error for the spatial-correlation function. When the noise amount changes from 10% to 100%, the accuracy of our results almost does not change at all. This convincingly shows that our extraction algorithm is very resilient to the measurement noise.

We further test the robustness of our algorithm by reducing the number of chip samples N from 2000 to 1500, 1000, and 500. We see that when there are reasonable number of chip samples (1500 and 1000), our algorithm still gives quite accurate results, and the maximum error for the global variation is no more than 10%, and the maximum error in either the spatial variation or spatial-correlation function is less than 5%. When the chip samples drop to 500, we start to see a larger error (but no more than 20%) in the extracted global variation. These observations are expected, because according to the statistical sampling theories, there is a lower bound on the number of samples in order to obtain reasonably accurate statistics.

Moreover, we observe that because of the optimization procedure used to extract the spatial variation and spatial-correlation function, as shown in (35), the extraction of those two parts is not as sensitive to the number of sample chips as the global-variation extraction does.

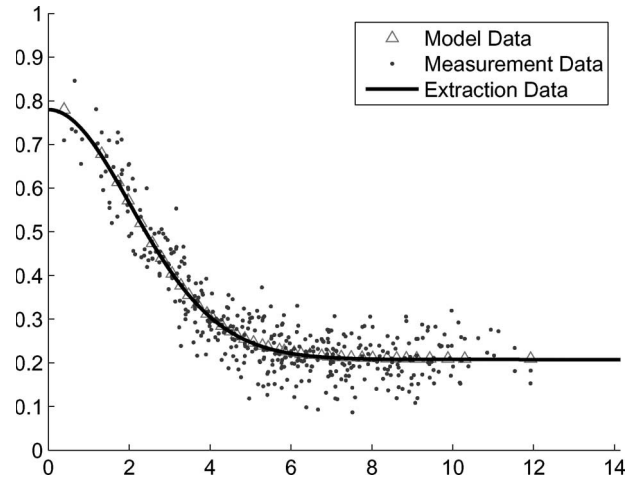

Fig. 9. Experiment on extracting the overall process correlation function.

We further fix the number of sample chips N to be 1000 and vary the number of measurement sites M on the chip from 60 to 50 and 40 to study how the accuracy of our algorithm changes. From Table I, we see that our algorithm still gives quite accurate results. When M changes from 60 to 40, we only see slight increase of errors for all extracted variation components, and none of them has more than 10% error.

We further plot one of the extracted overall process correlation functions in Fig. 9, where the (red) triangle points are the model data from the Monte Carlo model and the (blue) dotted points are data from our measurements with noise added. Obviously, the measurement data are noisy, not consistent, and are quite difficult to use directly. But after applying our algorithm, we obtain a very robust yet consistent results as shown in the (black) continuous curve, which not only captures the underlying process model, but also provide consistent extrapolation results for those data points that are not even available from measurement.

B. Extraction of Valid Spatial-Correlation Matrix

In the second experiment, we obtain the measurement data for M number of grids of interest on the chip based on a Monte Carlo model with some intentionally generated inconsistency (either through missing measurement data or adding Gaussian noise). We want to obtain the overall process correlation matrix for the M number of grids. We apply the algorithm, as shown in Fig. 7, to achieve this goal.

We show experimental results in Table II. According to the algorithm, as shown in Fig. 7, we compute individual pairwise correlations and, then, put them together to obtain an estimated correlation matrix A. Because of measurement noise, the resulting correlation matrix may not be positive semidefinite as illustrated by the second row in Table II, where the smallest eigenvalue λ_{least} of A is shown. For example, when we have 200 points, the measured correlation matrix has the smallest eigenvalue -2.38. The negative eigenvalue indicates that the measured correlation matrix is not positive semidefinite. On the contrary, after applying the modified alternative project algorithm, as shown in Fig. 8, we can always find a "closest"

TABLE II
OVERALL PROCESS-CORRELATION-MATRIX EXTRACTION

Points	50	100	150	200
$\lambda(A)_{least}$	-0.83	-1.43	-1.84	-2.38
$\lambda(\Omega)_{least}$	0	0	0	0
$\| A-\Omega \|_2$	2.09	4.35	6.85	9.39
$\frac{\|A-\Omega\|_2}{\|A\|_2}$	5.2%	5.9%	6.6%	7.3%
Iterations	23	34	38	41

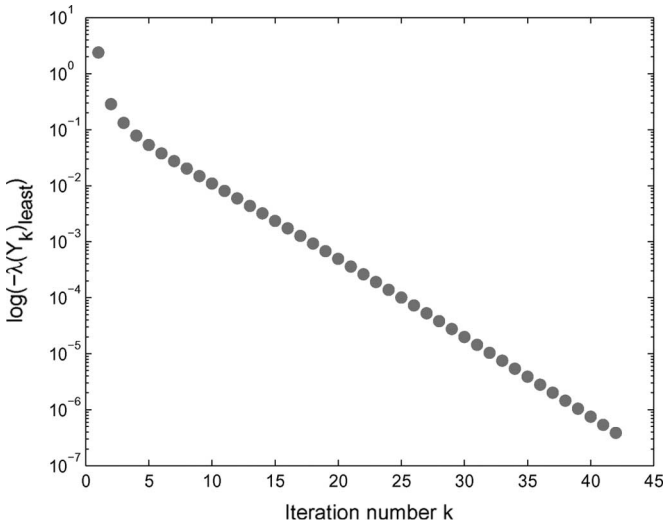

Fig. 10. Change of the least eigenvalue of Y_k in the alternative-projection algorithm, as shown in Fig. 7.

yet valid correlation matrix Ω. Moreover, the resulting matrix Ω has all nonnegative eigenvalues, as shown in the third row in Table II. Moreover, the difference between Ω and A is very small (no more than 10%).

We also report the number of iterations needed for the algorithm, as shown in Fig. 7, to converge. We find that the algorithm converges reasonable fast, and it takes only 41 iterations for the largest test case with $M = 200$. We further plot the change of the least eigenvalue of Y_k (which is negative) in each iteration in Fig. 10, where the y axis is the log-plot of the negative of the least eigenvalue of Y_k, and y axis is the iteration numbers. According to Fig. 10, we see that the least eigenvalue of Y_k are improved quickly in the first two iterations, and then its improvement becomes relative stable in the following iterations.

According to the definition of rate of convergence [28], we have

$$\lim_k \frac{|\lambda(Y_{k+1})_{\text{least}}|}{|\lambda(Y_k)_{\text{least}}|} = \mu \tag{53}$$

where the number μ is called the rate of convergence, and it should be between zero and one. If $\mu = 0$, then the sequence of $\lambda(Y_k)_{\text{least}}$ converges superlinearly. Otherwise, the sequence converges linearly with the rate of convergence of μ. We plot the estimated rate of convergence in each iteration in Fig. 11. We observe that the alternative-projection algorithm indeed has a linear convergence, which is in agreement with the theoretical results given by [26]. The rate of convergence in this particular example as shown in Fig. 11 is approximately 0.72.

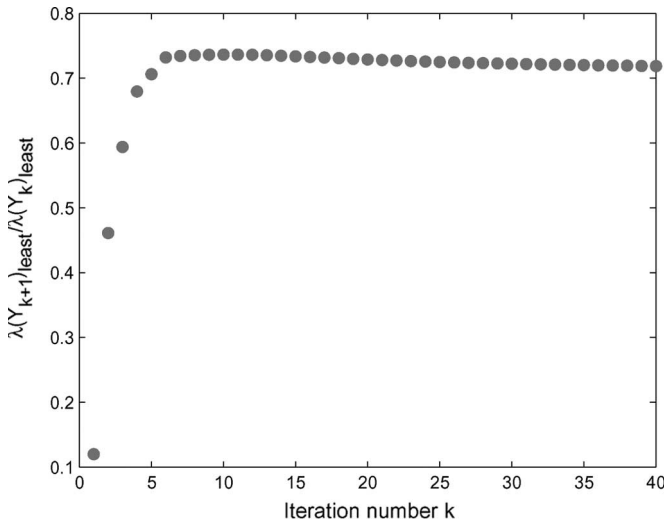

Fig. 11. Estimated rate of convergence of the alternative-projection algorithm, as shown in Fig. 7.

In summary, our experiment results convincingly show that our proposed extraction algorithms can accurately extract different variation components and are robust to the unavoidable measurement noise. Moreover, it is guaranteed that our algorithms always produce a valid spatial-correlation function or spatial-correlation matrix, which warrants the validity of further operations on these extracted variation data.

VII. Conclusion and Discussion

Robust extraction of statistical characteristics of process parameters is essential to achieve the benefits provided by statistical timing analysis and robust circuit optimization. In this paper, we have developed a novel technique to robustly extract the statistical characteristics of process variation from experimental measurements. Our technique guarantees that the resulting spatial-correlation function and spatial-correlation matrix are always valid and are the closest to the measurement data even if the data are inconsistent or distorted by some measurement noise.

In this paper, we have assumed that the spatial correlation follows a Gaussian random process, hence, only second-order moments are enough to characterize the variation. In the future, we will remove such an assumption and develop techniques that apply to more general random processes. We also plan to apply this technique to real wafer data and use the extracted process characteristics for robust mixed-signal-circuits tuning with consideration of correlated process variations in the future.

Acknowledgment

The authors would like to thank Dr. C. Visweswariah at the IBM Thomas J. Watson Research Center for his helpful discussion and suggestions in this paper. The authors would also like to thank the anonymous reviewers for their review comments, which helped to improve the presentation of this paper significantly.

References

[1] H. Chang and S. S. Sapatnekar, "Statistical timing analysis considering spatial correlations using a single PERT-like traversal," in *Proc. Int. Conf. Comput.-Aided Design*, Nov. 2003, pp. 621–625.

[2] C. Visweswariah, K. Ravindran, K. Kalafala, S. Walker, and S. Narayan, "First-order incremental block-based statistical timing analysis," in *Proc. Des. Autom. Conf.*, Jun. 2004, pp. 331–336.

[3] M. Orshansky and A. Bandyopadhyay, "Fast statistical timing analysis handling arbitrary delay correlations," in *Proc. Des. Autom. Conf.*, Jun. 2004, pp. 337–342.

[4] A. Agarwal, D. Blaauw, and V. Zolotov, "Statistical timing analysis for intra-die process variations with spatial correlations," in *Proc. Int. Conf. Comput.-Aided Des.*, Nov. 2003, pp. 900–907.

[5] L. Zhang, W. Chen, Y. Hu, J. A. Gubner, and C. C.-P. Chen, "Statistical timing analysis with extended pseudo-canonical timing model," in *Proc. Des. Autom. Test Eur.*, Mar. 2005, pp. 952–957.

[6] J. Kibarian and A. Strojwas, "Using spatial information to analyze correlations between test structure data," *IEEE Trans. Semicond. Manuf.*, vol. 4, no. 3, pp. 219–225, Aug. 1991.

[7] P. Mozumder and L. Loewenstein, "Method for semiconductor process optimization using functional representations of spatial variations and selectivity," *IEEE Trans. Compon., Hybrids, Manuf. Technol.*, vol. 15, no. 3, pp. 311–316, Jun. 1992.

[8] B. Stine, D. Boning, and J. Chung, "Analysis and decomposition of spatial variation in integrated circuit processes and devices," *IEEE Trans. Semicond. Manuf.*, vol. 10, no. 1, pp. 24–41, Feb. 1997.

[9] M. Orshansky, L. Milor, P. Chen, K. Keutzer, and C. Hu, "Impact of spatial intrachip gate length variability on the performance of high-speed digital circuits," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 21, no. 5, pp. 544–553, May 2002.

[10] J.-S. Doh, D.-W. Kim, S.-H. Lee, J.-B. Lee, Y.-k. Park, M.-H. Yoo, and J.-T. Kong, "A unified statistical model for inter-die and intra-die process variation," in *Proc. Int. Conf. SISPAD*, Sep. 2005, pp. 131–134.

[11] P. Friedberg, Y. Cao, J. Cain, R. Wang, J. Rabaey, and C. Spanos, "Modeling within-die spatial correlation effects for process-design co-optimization," in *Proc. 6th Int. Symp. Quality Electron. Des.*, Mar. 2005, pp. 516–521.

[12] J. Xiong, V. Zolotov, and L. He, "Robust extraction of spatial correlation," in *Proc. Int. Symp. Phys. Des.*, Apr. 2006, pp. 2–9.

[13] S. G. Duvall, "Statistical circuit modeling and optimization," in *Proc. 5th Int. Workshop Statistical Metrology*, Jun. 2000, pp. 56–63.

[14] S. Nassif, "Modeling and analysis of manufacturing variations," in *Proc. IEEE Int. Conf. Custom Integr. Circuits*, May 2001, pp. 223–228.

[15] K. A. Bowman, S. G. Duvall, and J. D. Meindl, "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *IEEE J. Solid-State Circuits*, vol. 37, no. 2, pp. 183–190, Feb. 2002.

[16] S. Nassif, D. Boning, and N. Hakim, "The care and feeding of your statistical static timer," in *Proc. Int. Conf. Comput.-Aided Des.*, Nov. 2004, pp. 138–139.

[17] P. Gupta and F.-L. Heng, "Toward a systematic-variation aware timing methodology," in *Proc. Des. Autom. Conf.*, Jun. 2004, pp. 321–326.

[18] L. Chen, L. Milor, C. Ouyang, W. Maly, and Y. Peng, "Analysis of the impact of proximity correction algorithms on circuit performance," *IEEE Trans. Semicond. Manuf.*, vol. 12, no. 3, pp. 313–322, Aug. 1999.

[19] A. B. Kahng and Y. C. Pati, "Subwavelength lithography and its potential impact on design and EDA," in *Proc. Des. Autom. Conf.*, Jun. 1999, pp. 799–804.

[20] R. Hogg and E. Tanis, *Probability and Statistical Inference*. Englewood Cliffs, NJ: Prentice-Hall, 2001.

[21] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra.* Philadelphia, PA: SIAM, 2001.

[22] A. Yaglom, "Some classes of random fields in n-dimensional space, related to stationary random processes," *Theory Probab. Appl.*, vol. 2, no. 3, pp. 273–320, 1957.

[23] R. L. Bras and I. Rodriguez-Iturbe, *Random Functions and Hydrology*. New York: Dover, 1985.

[24] M. Armstrong and R. Jabin, "Variogram models must be positive-definite," *Math. Geol.*, vol. 13, no. 5, pp. 455–459, Oct. 1981.

[25] T. Coleman and Y. Li, "An interior, trust region approach for nonlinear minimization subject to bounds," *SIAM J. Optim.*, vol. 6, no. 2, pp. 418–445, 1996.

[26] N. Higham, "Computing the nearest correlation matrix—A problem from finance," *IMA J. Numer. Anal.*, vol. 22, no. 3, pp. 329–343, 2002.

[27] S. Boyd and L. Xiao, "Least-squares covariance matrix adjustment," *SIAM J. Matrix Anal. Appl.*, vol. 27, no. 2, pp. 532–546, 2005.

[28] M. Schatzman, *Numerical Analysis: A Mathematical Introduction*. Oxford, U.K.: Clarendon, 2002.

Jinjun Xiong (S'05–M'07) received the B.E. degree (with honors) in precision instrument, the B.E. degree in industrial engineering, and the M.E. degree in precision instrument from Tsinghua University, Beijing, China, in 1998, 1998, and 2000, respectively, the M.S. degree in electrical and computer engineering from the University of Wisconsin, Madison, in 2001 and the Ph.D. degree in electrical engineering from the University of California (UCLA), Los Angeles, in 2006.

He is currently a Research Staff member at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY. His research interests include statistical timing analysis and optimization, design for manufacturability, design automation for very large-scale integrated circuits and systems, large-scale optimization and combinatorial mathematics.

Dr. Xiong is the recipient of the Distinguished Graduate Fellowship from the University of Wisconsin, Madison, in 2001, and from the UCLA, in 2002. He is the recipient of the Best Student Paper Award at the International Conference, on ASIC 2003, and the Best Paper Award at the ACM International Symposium on Physical Design, in 2006. He is also the recipient of the 2005–2006 Outstanding Ph.D. Award in electrical engineering from the UCLA.

Vladimir Zolotov (M'97–SM'04) received the M.S. degree on electrical engineering from Moscow Institute of Electronics, Moscow, Russia, and the Ph.D. degree on electrical engineering from Scientific Research Institute of Micro Devices, Moscow.

He is currently a Research Staff member of IBM T. J. Watson Research Center at Yorktown Heights, NY, where he is working on statistical static-timing analysis. Previously, he was with Motorola Inc., where he was involved in the development of electronic design-automation tools and methodology for high-performance and low-power very large-scale integration (VLSI) designs. His research interests include statistical timing, signal integrity, fast circuit simulation, reliability analysis, onchip inductance, and optimization of VLSI.

Lei He (S'94–M'99) received the Ph.D. degree in computer science from the University of California (UCLA), Los Angeles, in 1999.

He is currently an Associate Professor with the Electrical Engineering Department, UCLA, and was a faculty member at University of Wisconsin, Madison, between 1999 and 2001. He also held visiting or consulting positions with Intel, Hewlett-Package, Cadence, and Synopsys. His research interests include very large-scale integrated circuits and systems and electronic-design automation. He has published over 130 technical papers and is a technical program committee member for a number of conferences including Design Automation Conference, International Conference on Computer-Aided Design, International Symposium on Low-Power Electronics and Design, and International Symposium on Field Programmable Gate Array.

Dr. He was the recipient of a U.S. National Science Foundation CAREER Award in 2000, a UCLA Chancellor's Faculty Career Development Award (highest class) in 2003, an IBM Faculty Award in 2003, a Northrop Grumman Excellence in Teaching Award in 2005, and a Best Paper Award in the 2006 International Symposium on Physical Design.