Robust Extraction of Spatial Correlation *

Jinjun Xiong
EE Dept., Univ. of California
Los Angeles, CA 90095
jinjun@ee.ucla.edu

Vladimir Zolotov
IBM Research Center
Yorktown Heights, NY 10598
zolotov@us.ibm.com

Lei He
EE Dept., Univ. of California
Los Angeles, CA 90095
lhe@ee.ucla.edu

ABSTRACT

Increased variability of process parameters and recent progress in statistical static timing analysis make extraction of statistical characteristics of process variation and spatial correlation an important yet challenging problem in modern chip designs. Unfortunately, existing approaches either focus on extraction of only a deterministic component of spatial variation or do not consider actual difficulties in computing a valid spatial correlation function and matrix, simply ignoring the fact that not every function and matrix can be used to describe the spatial correlation. Based upon the mathematical theory of random fields and convex analysis, in this paper, 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 that are closest to measurement data, even if those measurements are affected by unavoidable random noises. Experiment results based upon 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.

Categories and Subject Descriptors

J.6 [**COMPUTER-AIDED ENGINEERING**]: Computer-aided design (CAD)

General Terms

Algorithms, Measurement, Design, Reliability.

*This work is supported in part by the NSF CAREER award CCR-0401682, a UC MICRO grant sponsored by Actel, and a Faculty Partner Award by IBM.

ISPD'06, April 9–12, 2006, San Jose, California, USA.

Keywords

Spatial Correlation, Extraction, Process Variation, Robust Extraction.

1. INTRODUCTION

Aggressive scaling down of transistors and interconnects has resulted in miraculous achievements in chip performance and functionality. However, this deep scaling of semiconductor technology has introduced the problem of uncontrollable process variations. Nowadays, it is commonly recognized that for deep submicron technologies we are simply unable to make transistors and interconnects with accurately predictable characteristics. Moreover, we are not even able 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 approach is supported by the successful development of statistical static timing analysis (SSTA) tools capable of predicting statistical timing yield [1, 2, 3] of designed chips. Modern SSTA tools can handle both inter-chip and intra-chip random variations of process and environmental parameters.

The inter-chip variations represent global variations that are the same for all devices on a given chip. The intra-chip variations represent variations of devices within the same chip. The intra-chip variations include spatially correlated variations and purely independent or uncorrelated variations. We say that parameter F has spatially correlated variation if devices that are closer to each other are more likely to have similar values of this parameter, while devices that are far away are more likely to have different values of the parameter.

It is of significance to characterize different process variation components (global, spatial and random variations) and the overall process correlation, because that information is the enabler 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 essential inputs for both statistical timing analysis and robust circuit tuning. Recent statistical static timing analysis techniques considering spatially correlated parameters [1, 4, 5] also assume that the required spatially correlated information is known *a priori*. In fact, the only way to obtain these variation characteristics is to extract them from experimental measurements. However, to the best of our knowledge, no existing work has provided a detailed

technique to extract that information properly from measurements. Moreover, how to robustly extract statistical characteristics of device parameters and especially characteristics of spatially correlated parameters in the presence of unavoidable measurement noise has not been considered in literature [6, 7, 8]. The extraction of the deterministic component of Leff variation was considered in detail in [9] for the 0.18μm CMOS technology. But that publication ignored the random component of spatial variations, justifying its approach by the fact that for the 0.18μm CMOS technology random variations were insignificant. Another publication [10] limited its consideration by simple computation of the spatial correlation coefficient that is a function of distance. However, there is no verification that the extracted correlation function was a valid correlation function, i.e., any correlation matrix generated from this function must be positive semidefinite.

From the theory of random fields [11], it is known that not any function can be a valid spatial correlation function. Even not every monotone decreasing function can represent spatial correlation because it cannot guarantee to generate a valid spatial correlation matrix for any given set of locations on a chip. Moreover, not every set of coefficients can represent a valid correlation matrix even if all those coefficients are less than one, because the resulting matrix simply can be non-positive semidefinite. On the one hand, because of unavoidable measurement errors, it is likely to extract a set of correlation coefficients that do not form a valid correlation matrix or cannot be interpolated into a valid correlation function. On the other hand, statistical static timing analysis requires that the correlation function is valid. Otherwise, the results of statistical timing can have unpredictable error.

The major contribution of this work is as follows. 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 theory 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 to the measured data 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.

The rest of the paper is organized as follows. We first describe how to model process variations in section 2; then present our problem formulations in section 3; and provide algorithms to solve the problems in section 4 and 5, respectively. Experiment results are presented in section 6, and we draw conclusions in section 7.

2. MODELING OF PROCESS VARIATION

2.1 Process Variation Decomposition

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, like gate delay and threshold voltage. We model the parameter as a random variable, and its overall variation can be decomposed into three distinct components: the inter-chip global variation X, the intra-chip spatial variation Y, and the purely uncorrelated random variation Z, i.e.,

$$F = f_0 + X + Y + Z, \quad (1)$$

where f_0 is the mean value of F, and X, Y and Z are random variables. Studies have shown that the mean value f_0 may also exhibit *systematic variation*. How to extract the systematic variation has been studied in detail in [9]. Therefore, in the following we do not consider systematic variation in our model. Instead, we focus on the zero-mean random variation components X, Y and Z. The inter-chip global variation X models the variation due to global variation effects that are shared for all device parameters within the chip, hence it will be same for all measurements within the same chip but may be different for different chips. The intra-chip spatial variation Y models location-dependent variations within the chip, hence it may be different for different measurements at different locations within the same chip. The random variation Z models the purely uncorrelated random component that is not explainable by either the inter-chip global variation X or intra-chip spatial variation Y. Therefore, X, Y and Z have mutually independent distributions, as the mechanisms of causing variation in X, Y and Z are different by definition. The variance of the parameter of interest F is given by

$$\sigma_F^2 = \sigma_X^2 + \sigma_Y^2 + \sigma_Z^2, \quad (2)$$

where σ_X^2, σ_Y^2, and σ_Z^2 are the variances of X, Y, and Z, respectively. The total variance σ_F^2 is also called the *overall chip variance*.

2.2 Modeling of Spatial Correlation

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 order to model the spatial correlation of random variable Y, we associate every point (x, y) in a chip with a random variable $Y(x, y)$. We have the following definitions:

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

DEFINITION 2. **Homogeneous and Isotropic Random Field** *is a random field $Y(x, y)$ whose mean value is a constant number, and whose correlation function $\rho(x_i, x_j, y_i, y_j)$ between any two points depends only on the distance v between them, i.e.,*

$$\rho(x_i, x_j, y_i, y_j) = \rho(\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}) = \rho(v). \quad (3)$$

For any number of chosen points on the chip, we assume the joint spatial variation $\mathbf{Y}=(Y_1, Y_2, ..., Y_M)^T$ follows a multivariate Gaussian process with respect to their respective physical locations on the chip. Therefore, to fully characterize the M-dimensional Gaussian distribution, we only need to know the variance σ_Y^2 and its correlation matrix Ω as

shown in (4):

$$\Omega = \begin{bmatrix} 1 & \rho_{1,2} & \rho_{1,3} & ... & \rho_{1,M} \\ \rho_{1,2} & 1 & \rho_{2,3} & ... & \rho_{2,M} \\ \rho_{1,3} & \rho_{2,3} & 1 & ... & \rho_{3,M} \\ \cdot & & & & \cdot \\ \rho_{1,M} & \rho_{2,M} & \rho_{3,M} & ... & 1 \end{bmatrix}. \quad (4)$$

A *valid* correlation matrix must be *positive semidefinite* [12].

If the spatial variation is modeled as a *homogeneous and isotropic random field* in a two dimensional space $\mathcal{R}^2$, we can also characterize it by extracting its variance σ_Y^2 and the correlation function $\rho(v)$.

For the parameter of interest at two different points, their covariance is

$$\begin{aligned} cov(F_i, F_j) &= cov(X + Y_i + Z, X + Y_j + Z) \\ &= cov(X, X) + cov(Y_i, Y_j) \\ &= \sigma_X^2 + \rho(v_{i,j})\sigma_Y^2, \end{aligned} \quad (5)$$

where $\rho(v_{i,j})$ is the spatial correlation coefficient between two locations that are $v_{i,j}$ distance apart. Saying that the spatial variation follows a homogeneous and isotropic random field is equivalent to saying that 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.

In general, the spatial correlation function $\rho(v)$ is a monotonically decreasing function of distance v. However, not all monotonically decreasing functions qualify for the spatial correlation function. We have 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*

$$\rho(v) = \int_0^\infty J_0(\omega v) d(\Phi(\omega)), \quad (6)$$

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 non-negative p,

$$\int_0^\infty \frac{d\Phi(\omega)}{(1+\omega^2)^p} < \infty. \quad (7)$$

Proof: See [11] for the proof. □

For example, $\Phi(\omega)$ =1-$\frac{1}{\sqrt{1+\omega^2/b^2}}$ and $\Phi(\omega)$=1-$exp(-\frac{\omega^2}{4b^2})$ satisfy this condition. By plugging them into (6), we obtain the corresponding correlation functions as $\rho(v)$=$exp(-bv)$ and $\rho(v)$=$exp(-b^2v^2)$, respectively, where b is a parameter that regulates the decaying rate of the correlation function with respect to distance v.

3. PROBLEM FORMULATION

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 k^{th} chip on the i^{th} site.

When the spatial variation follows a homogeneous and isotropic random field, we have the following first problem formulation:

FORMULATION 1. **Extraction of Process Variation**: *Given noisy measurement data for the parameter of interest with possible inconsistency, extract the inter-chip global variation component σ_X^2, the intra-chip spatial variation component σ_Y^2, the random variation component σ_Z^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 (6).*

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

$$\rho_v \equiv \frac{cov(F_i, F_j)}{\sigma_{F_i}\sigma_{F_j}} \quad (8)$$

$$= \frac{\sigma_X^2 + \rho(v)\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2}. \quad (9)$$

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.

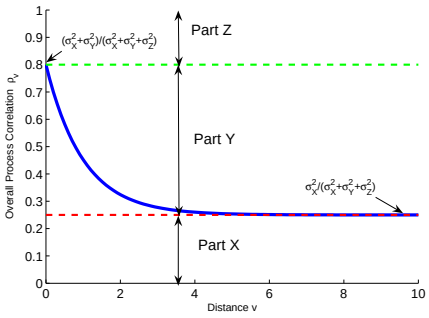

Figure 1: Overall process correlation.

In Figure 1, we show a possible curve for the overall correlation ρ_v as a function of the distance v. According to Figure 1, the total correlation can be divided into three parts: part X is the correlation caused by the inter-chip global variation; part Y is the correlation caused by the intra-chip spatial correlation; and part Z 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, 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.

In some applications, however, it may not be appropriate to model the spatial correlation as a homogeneous and isotropic random field. In this case, we have to use the spa-

tial correlation matrix to characterize the spatial variation directly. Therefore, we have the following second problem formulation:

FORMULATION 2. **Extraction of Spatial Correlation Matrix**: *Given noisy measurement data for the parameter of interest with possible inconsistency at M number of points on a chip, extract the overall process spatial correlation matrix Ω as shown in (4), so that it not only accurately captures the underlying process variation model, but the extracted correlation matrix Ω is always positive semidefinite.*

In other words, we want to characterize the overall process spatial correlation by providing the symmetric correlation matrix Ω directly as shown in (4) for the given M number of points of interest. Extracting a valid correlation matrix is of practical use. For example, in a PCA (principle component analysis) based statistical timing analysis [1], the required spatial correlation matrix is always assumed to be valid and known *a priori*.

In the following, we present techniques to solve the above two problem formulations respectively.

4. EXTRACTION OF PROCESS VARIATION

4.1 Global Variation Extraction

We treat each measurement of the parameter of interest F as a sampling of the quantity in (1). We also group the measurement data $f_{k,i}$ by their chip locations as follows: $f_{k,.}$=$[f_{k,1}, ..., f_{k,M}]$ for k=1 to N, or by their site locations as follows: $f_{.,i}$=$[f_{1,i}, ..., f_{N,i}]$ for i=1 to M. For better presentation, we denote the actual variance as σ^2 with an upper case letter in subscript, like σ_X^2 for the global variation component; and denote the extracted variance as σ^2 with a lower case letter in subscript, like σ_x^2 for the extracted global variation component.

We approximate the overall chip variance σ_F^2 by computing the *unbiased sample variance* [12] of $f_{k,i}$ as follows

$$\sigma_F^2 \approx \sigma_f^2 = \frac{1}{MN-1}(\sum_i \sum_k f_{k,i}^2 - \frac{(\sum_i \sum_k f_{k,i})^2}{MN}). \quad (10)$$

For all samples of the parameter of interest F within a particular chip c, because the inter-chip global variation X 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

$$\sigma_{F_c}^2 \quad = \quad \sigma_Y^2 + \sigma_Z^2. \quad (11)$$

We estimate the overall within-chip variation by computing the *unbiased sample variance* [12] of $f_{k,.}$ as follows

$$\sigma_{F_c}^2 \approx \sigma_{f_k}^2 = \frac{1}{M-1}(\sum_i f_{k,i}^2 - \frac{(\sum_i f_{k,i})^2}{M}). \quad (12)$$

For different $f_{k,.}$, we may get different estimation of $\sigma_{F_c}^2$ due to possibly 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 inter-chip global variation by

$$\sigma_X^2 = \sigma_F^2 - \sigma_{F_c}^2 \approx \sigma_x^2 = \sigma_f^2 - \sigma_{f_c}^2. \quad (13)$$

4.2 Spatial Variation Extraction

For any two different sets of $f_{.,i}$ and $f_{.,j}$ 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* [12] of $f_{.,i}$ and $f_{.,j}$ as follows

$$\begin{aligned} cov(F_i, F_j) &\approx cov(f_{.,i}, f_{.,j}) \\ &= \frac{\sum_k f_{k,i} f_{k,j}}{N-1} - \frac{\sum_k f_{k,i} \sum_k f_{k,j}}{N(N-1)}. \end{aligned} \quad (14)$$

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

According to (5) and (13), we can estimate the product of spatial variation σ_Y^2 and spatial correlation $\rho(v)$ as follows:

$$\sigma_Y^2 \rho(v) = cov(F_i, F_j) - \sigma_X^2 \approx cov(v) - \sigma_x^2. \quad (15)$$

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 $\rho(v)$. However, there are two challenges in doing that: (1) we do not know the exact value of spatial variation σ_Y^2; (2) due to 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_Y^2 accurately. Moreover, the resulting $\rho(v)$ is guaranteed to be a valid spatial correlation function for modeling the homogeneous and isotropic random field, and any spatial correlation matrix generated from $\rho(v)$ is guaranteed to be positive semidefinite.

Given the data set (v, $cov(v)$) as computed from (14), we formulate the robust spatial variation extraction problem as the following optimization problem:

$$\begin{aligned} \min_{\Phi, \sigma_Y^2} : \quad & \| \sigma_Y^2 \int_0^\infty J_0(\omega v) d(\Phi(\omega)) - cov(v) + \sigma_x^2 \|, \quad (16) \\ s.t. \quad & \sigma_Y^2 \le \sigma_{f_c}^2, \\ & \int_0^\infty \frac{d\Phi(\omega)}{(1+\omega^2)^p} < \infty. \end{aligned}$$

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 (6) to obtain the valid spatial correlation function $\rho(v)$.

The above problem formulation is very general and applies to any real non-decreasing function $\Phi(\omega)$. However, for practical use, there is no need to enumerate all possible choices of $\Phi(\omega)$ in order to find the optimal $\rho(v)$. In fact, it is sufficient to chose a family of functions $\Phi(\omega)$ so that the $\rho(v)$ obtained from (6) contains a rich set of functions for the purpose of modeling spatial correlation.

It has been shown in [13] that by choosing a proper family function of $\Phi(\omega)$, we can obtain a very general family of spatial correlation functions

$$\rho(v) = 2(\frac{bv}{2})^{s-1} K_{s-1}(bv) \Gamma(s-1)^{-1}, \quad (17)$$

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 can obtain different spatial correlation functions.

For example, $\rho(v)$=$exp(-bv)$ with $\Phi(\omega)$ =1-$\frac{1}{\sqrt{1+\omega^2/b^2}}$ in (6) can be generated from (17) by choosing $s = 3/2$.

To show that the function of (17) indeed provides us a rich set of correlation functions that suffice for our spatial correlation modeling, we plot the function of (17) under different parameters of b and s. Figure 2 shows a few samples of correlation functions generated from (17) by setting b to be 0.1, 1 and 10, and varying s from 2 to 10 with a step size of 2. From the figure, we can see that the correlation function (17) indeed can generate a rich gamut of correlation functions for the purpose of spatial correlation modeling.

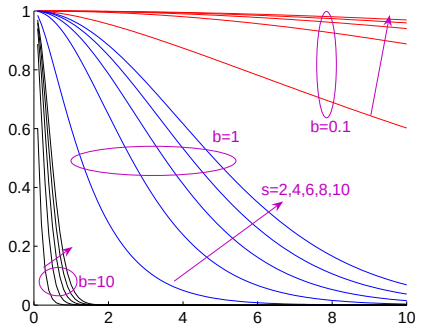

Figure 2: Correlation functions generated from (17).

Without loss of generality, in the following, equation (17) will be used as the candidate correlation function in (16). Moreover, 2-norm is used as a measure of the objective function in (16). Therefore, we can rewrite the optimization problem as given in (16) as follows:

$$\begin{aligned} \min_{b,s,\sigma_Y^2} : \quad & \sum [2\sigma_Y^2 (\frac{bv}{2})^{s-1} K_{s-1}(bv)\Gamma(s-1)^{-1} \\ & -cov(v) + \sigma_x^2]^2, \\ s.t. \quad & \sigma_Y^2 \leq \sigma_{f_c}^2. \end{aligned} \tag{18}$$

This is a constrained nonlinear least square problem and we can solve it efficiently via any nonlinear least square technique [14].

After solving the above problem, we obtain the estimated spatial variation component $\sigma_y^2 \approx \sigma_Y^2$, and the parameter b and s. By plugging b and s into (17), 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.

4.3 Overall Algorithm

The overall algorithm for characterizing the process variation is summarized as shown in Figure 3:

1	Extract global variation σ_x^2 by (13);
2	Solve (18) to obtain σ_y^2 and b and s;
3	Extract $\overline{\rho(v)}$ by plugging b and s into (17);
4	Extract random variation σ_z^2 by (19);
5	Extract overall process correlation by (9);

Figure 3: Algorithm for characterization of process variation.

We first extract the global variation component σ_x^2 by using formula (13). We then solve the nonlinear least square optimization problem as defined in (18) to obtain the spatial variation component σ_y^2, and the parameter of b and s that define the spatial correlation function for a homogeneous and isotropic random field as shown in (17). According to (2), we extract the random variation component by using the following formula:

$$\sigma_Z^2 = \sigma_F^2 - \sigma_X^2 - \sigma_Y^2 \approx \sigma_z^2 = \sigma_f^2 - \sigma_x^2 - \sigma_y^2. \tag{19}$$

By plugging all variation components into (9), we obtain the overall process correlation at any distance.

5. EXTRACTION OF SPATIAL CORRELATION MATRIX

5.1 Overall Algorithm

We are given measurement data for M points of interest on a chip and we have N samples of the same chip. We extract the overall process spatial correlation as follows.

We first estimate the covariance between any two points of interest by (14). We then estimate the variance of each point of interest by computing its *unbiased sample variance* [12] as follows:

$$\sigma_{F_i}^2 \approx \sigma_{f_i}^2 = \frac{1}{N-1}(\sum_k f_{k,i}^2 - \frac{(\sum_k f_{k,i})^2}{N}). \tag{20}$$

By plugging the estimated $\sigma_{f_i}^2$ and $\sigma_{f_j}^2$ and $cov(f_{.,i}, f_{.,j})$ from (14) into (8), we obtain the estimated overall process correlation coefficient

$$\rho_{i,j} = \frac{cov(F_i, F_j)}{\sigma_{F_i}\sigma_{F_j}} \approx \frac{cov(f_{.,i}, f_{.,j})}{\sigma_{f_i}\sigma_{f_j}}. \tag{21}$$

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 $\rho_{i,j}$. Putting all $\rho_{i,j}$ into (4), 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 due to measurement errors, we can not guarantee that such a property would hold automatically for the resulting A due to the unreliable (sometimes even inconsistent) measurement data. We solve this problem by employing the *modified alternative projection algorithm* that enables us 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 Figure 4:

1	Compute $cov(f_i, f_j)$ by (14);
2	Compute $\sigma_{f_i}^2$ by (20);
3	Compute $\rho_{i,j}$ by (21);
4	Compute A by assembling $\rho_{i,j}$ into (4);
5	Compute Ω via the modified alternative projection algorithm;

Figure 4: Algorithm for extracting a valid spatial correlation matrix.

5.2 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 0 and 1, 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.,

$$\min_{\Omega} :\| A - \Omega \|, s.t. : \Omega \in correlation\ matrix. \quad (22)$$

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

$$\| A \|_W = \| W^{1/2} A W^{1/2} \|_F, \quad (23)$$

where W is a symmetric positive definite matrix.

This problem is also called the nearest correlation matrix problem [15], or the least-squares covariance adjustment problem [16]. We solve this problem by employing the modified alternative projection algorithm proposed in [15]. The idea is to iteratively project the symmetric matrix A onto two convex sets alternatively, and at the end of iteration, the final projected matrix is the solution to the optimization problem as defined in (22).

We first define the sets

$$U = \{Y = Y^T \in \mathcal{R}^{n \times n} : y_{ii} = 1\}, \quad (24)$$

$$S = \{Y = Y^T \in \mathcal{R}^{n \times n} : Y \geq 0\}, \quad (25)$$

where the notation $Y \geq 0$ means that Y is positive semidefinite. Our desired correlation matrix Ω as shown in (22) is a matrix 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 (22) is obtainable and is unique.

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

$$A_+ = Q diag(max(\lambda_i, 0)) Q^T. \quad (26)$$

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 the following formula.

$$P_U(A) = A - W^{-1} diag(\theta_i) W^{-1}, \quad (27)$$

where $\theta = [\theta_1, ..., \theta_n]^T$ is the solution of the linear system

$$(W^{-1} \circ W^{-1})\theta = diag(A - I), \quad (28)$$

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

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

$$P_S(A) = W^{-1/2}((W^{1/2} A W^{1/2})_+) W^{-1}. \quad (29)$$

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

$$P_U(A) = (p_{ij}) \quad (30)$$

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

$$P_S(A) = A_+ = Q diag(max(\lambda_i, 0)) Q^T. \quad (31)$$

The following modified alternative projection algorithm as shown in Figure 5 can be used to solve the nearest correlation matrix problem as defined in (22).

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

Figure 5: The modified alternative projection algorithm.

It has been proved that when $k \to \infty$, both X_k and Y_k converge to the desired correlation matrix Ω. Therefore, among many possible choices, the following convergence condition can be used in Figure 5 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 $\epsilon = 10^{-8}$).

6. EXPERIMENTAL RESULTS

We employ a Monte Carlo model of measurement to verify the robustness and accuracy of our extraction algorithms. The Monte Carlo model is based on a valid correlation function $\rho(v)$ with known variation amounts for all variation components. By comparing the extracted variation components σ_x^2, σ_y^2, and $\overline{\rho(v)}$ 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 wafter measurement.

In the first experiment, by using the Monte Carlo method, we generate 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 Figure 3, we extract the global variation component σ_x^2, random variation component σ_z^2, spatial variation component σ_y^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 (17). By plugging all variation components into (9), 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 $err(\sigma_X^2) = \frac{\sigma_x^2 - \sigma_X^2}{\sigma_X^2}$; for the spatial variation component, the relative error is given by $err(\sigma_Y^2) = \frac{\sigma_y^2 - \sigma_Y^2}{\sigma_Y^2}$; and for the spatial correlation function, the relative error is given by $err(\rho(v)) = \frac{\|\overline{\rho(v)} - \rho(v)\|}{\|\rho(v)\|}$.

From statistical theories, we know that if we have more measurement data, we have more confidence in the accuracy

of statistics obtained from measurements. However, in reality, 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 experiment results in Table 1, 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 ($\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2$). The product of N and M gives the total number of measurement points.

According to Table 1, 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.

Table 1: Process variation extraction.

N	M	$Noise$	$err(\sigma_X^2)$	$err(\sigma_Y^2)$	$err(\rho(v))$
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%

We further test the robustness of our algorithm by reducing the number of chip samples N from 2000 to 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. This is 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 (18), the extraction of those two parts is not as sensitive to the number of sample chips as the global variation extraction does.

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 40 to study how the accuracy of our algorithm changes. From Table 1, 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.

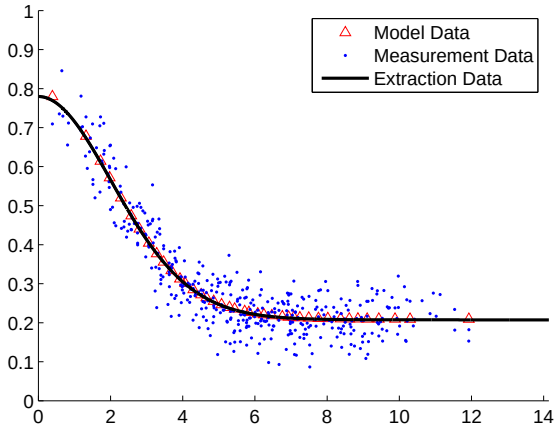

Figure 6: Experiment on extracting the overall process correlation function.

We further plot one of the extracted overall process correlation functions in Figure 6, where the (red) triangle points are the model data from the Monte Carlo model, 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. However, 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 very well, but also provide consistent extrapolation results for those data points that are not even available from measurement.

Table 2: 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
$\parallel A - \Omega \parallel_2$	2.09	4.35	6.85	9.39
$\frac{\parallel A-\Omega \parallel_2}{\parallel A \parallel_2}$	5.2%	5.9%	6.6%	7.3%

In the second experiment, we have measurement data for a number of points of interest on the chip, and we want to obtain their overall process correlation matrix for further operations. We apply the algorithm as shown in Figure 4 to achieve this goal.

We show experimental results in Table 2. According to the algorithm as shown in Figure 4, we compute individual pair-wise correlations and then put them together to obtain an estimated correlation matrix A. However, due to measurement noise, the resulting correlation matrix may not be positive semidefinite as illustrated by the second row in Table 2, 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 Figure 5, we can always find a "closest" yet valid correlation

matrix Ω. And the resulting matrix Ω has all non-negative eigenvalues as shown in the third row in Table 2. Moreover, the difference between Ω and A is very small (no more than 10%).

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.

7. CONCLUSION AND DISCUSSION

Robust extraction of statistical characteristics of process parameters is the enabler 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 distorted by some measurement noise. We 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.

8. REFERENCES

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

[2] C. Visweswariah, K. Ravindran, K. Kalafala, S. G. Walker, and S. Narayan, "First-order incremental block-based statistical timing analysis," in *Proc. Design Automation Conf*, June 2004.

[3] M. Orshansky and A. Bandyopadhyay, "Fast statistical timing analysis handling arbitrary delay correlations," in *DAC 04*, June 2004.

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

[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. Design Automation and Test in Europe*, pp. 952 – 957, March 2005.

[6] J. Kibarian and A. Strojwas, "Using spatial information to analyze correlations between test structure data," in *Semiconductor Manufacturing, IEEE Transactions on*, pp. 219 – 225, Aug. 1991.

[7] P. Mozumder and L. Loewenstein, "Method for semiconductor process optimization using functional representations of spatial variations and selectivity," in *Components, Hybrids, and Manufacturing Technology, IEEE Transactions on*, pp. 311 – 316, June 1992.

[8] B. Stine, D. Boning, and J. Chung, "Analysis and decomposition of spatial variation in integrated circuit processes and devices," in *Semiconductor Manufacturing, IEEE Transactions on*, 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," in *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, pp. 544–553, May 2002.

[10] 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 *Sixth International Symposium on Quality of Electronic Design*, pp. 516 – 521, March 2005.

[11] A. Yaglom, "Some classes of random fields in n-dimensional space, related to stationary random processes," in *Theory Probability Applications*, vol. 2, pp. 273–320, 1957.

[12] R. Hogg and E. Tanis, *Probability and Statistical Inference*. Prentice Hall, 2001.

[13] R. L. Bras and I. Rodriguez-Iturbe, *Random Funtions and Hydrology*. Dover Publishers, 1985.

[14] T. Coleman and Y. Li, "An interior, trust region approach for nonlinear minimization subject to bounds," in *SIAM Journal on Optimization*, vol. 6, pp. 418–445, 1996.

[15] N. Higham, "Computing the nearest correlation matrix - a problem from finance," in *IMA Journal of Numerical Analysis*, vol. 22, pp. 329–343, 2002.

[16] S. Boyd and L. Xiao, "Least-squares covariance matrix adjustment," in *SIAM Journal on Matrix Analysis and Applications*, June 2004. in submission.