Non-Parametric Approaches for Intra-Die Spatial Correlation Estimation

***Abstract*—TBD**

***Index Terms*—Spatial correlation, non-parametric estimation, anisotropic, B-spline, maximum likelihood estimation, concave-convex procedure.**

I. Introduction

A. Why spatial correlation is important?

As the minimum feature size of semiconductor device continues scaling down, integrated circuits suffer from increasing variations in the manufacturing process. These process variations lead to the geometric variations in the devices and interconnects, and greatly affect their electrical parameters. As a result, the performances of the fabricated circuits are degraded from the design specifications, and the manufacturing yield is lost. Thus, it is desirable to develop more accurate statistical analysis to tackle with variation problems in the design stages [**?**].

Process variations can be classified into two categories according to the spatial scales. *Inter-die variations* affect device parameters with the same value on a die but with different values beyond the die scope, while *intra-die variations* cause device parameter values to vary across different locations within a single die. As technology generation marches, intra-die variations exceed inter-die variations and become predominant in total process variations. Intra-die variations often show spatial correlated patterns, which means devices that are closely placed tend to possess similar characteristics than those further apart.

Spatial correlation is defined to describe the degree to which the values of device characteristics are related to each other as a function of spatial distances of the devices [**?**]. Assumed to be a given function or matrix form (what?), spatial correlation has been widely used in variation aware circuit analysis and design techniques, such as statistical timing analysis [**?**], [**?**], power/leakage minimization [**?**], [**?**].

B. Why anisotropic models?

In some circumstances, *isotropic* assumption is made for simplicity of computation. That is, the correlation is assumed to be only depend on the distance between two gates (or sites of random variables). However, based on the study of [?], certain variations such as critial dimension variation, exhibit significantly stronger correlation in horizontal direction than in verticle direction. In this paper, we focus on the anisotropic treatment of spatial correlation extraction.

C. Why non-parametric approaches make sense?

Moreover, in the previous researches, usually parametric approaches are used, i.e., the correlation function is parametrized in a simple form such as exponential function, Gaussian function or Matérn function. The advantages of this approach are that only a few parameters are need to be extracted, and the resulting correlation functions are guaranteed to be positive definite. However, those functions may not be able to accurately extract the actual correlation. Moreover, in [6], the authors observe that in parametric approaches, problems are expressed as a non-convex programming. The correlation matrix is expressed as a non-linear function of parameters. The numerical optimization solvers may get stuck in a local minimum. In non-parametric approach, correlation functions are curved-fitted using linear combination of some basis functions: $\rho(h) = \sum_i p_i \Phi_i(h)$, where p_i's are the unknown coefficients to be fitted and Φ_i's are a family of basis functions. The corresponding correlation matrix can be expressed as an affine function of the coefficients and hence convexity is preserved.

D. Why B-spline?

Besides the property of positive definiteness, practical correlation function may have other observable properties such as monotonicity and nonnegativity. The advantage of using B-spline is that the shapes of curves can be controlled easily be manipulating only the coefficients. Besides this, B-spline also has other nice properties such as ease of evaluation and locality.

E. Organization of this paper

In this paper, we will also discuss the difference between the least squares estimation and the maximum likelihood estimation. Problem forumation is given in Section **??**. Least squares estimation method and the maximum estimation method will be given in Section 3 and Section 4 respectively. Finally, conclusions and possible future works are given in Section 5.

II. Problem Formulation

According to [**?**], the intra-die variation Z can be further decomposed into three components:

- a *deterministic* component Z_{det}, which is determined by layout context and can be modeled by deliberatively exploring layout patterns;
- a *correlated random* (or *spatially correlated*) component Z_{cor}, which is random but shows correlated patterns due to proximity effects;

– a *purely random* component Z_{rnd}, which is spatially uncorrelated and can be treated as statistitically random.

In this paper, for the sake of simplicity, the deterministic component is assumed to be well modeled and taken away from the whole process variation. We concentrate on the spatially correlated component, together with the purely random component and measurement error as [**?**], [**?**]. The spatially correlated component is generally modeled as random field in the literature. In this section, fundamental concepts and theories of random field are reviewed. The purely random component and measurement error cause a discontinuity at the origin of the correlation function, which is called nugget effect. We will describe this phenomenon in this section.

A. Random Field [**?**]

Random field, also known as *stochastic process*, can be regarded as an indexed family of random variables denoted as $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$, where D is a subset of d-dimensional Euclidean space $\mathbb{R}^d$. To specify a stochastic process, the joint probability distribution function of any finite subset $(Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))$ must be given in a consistent way, which is called *distribution* of the process. For ease of analysis, a random field is often assumed to be with *Gaussian* distribution, and is called Gaussian random field.

A random field has several key properties that are useful in practical problems. The field is *stationary* under translations, or *homogeneous*, if the distribution is unchanged when the point set is translated. The field is *isotropic* if the distribution is invariant under any rotation of the whole points in the parameter space. We study homogeneous isotropic field in this paper.

The *covariance* C and *correlation* R of a stochastic process are defined by

$$C(\mathbf{s}_i, \mathbf{s}_j) = \text{cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = \text{E}[(Z(\mathbf{s}_i) - \text{E}[Z(\mathbf{s}_i)])(Z(\mathbf{s}_j) - \text{E}[Z(\mathbf{s}_j)])]$$

and

$$R(\mathbf{s}_i, \mathbf{s}_j) = C(\mathbf{s}_i, \mathbf{s}_j) / \sqrt{C(\mathbf{s}_i, \mathbf{s}_i) C(\mathbf{s}_j, \mathbf{s}_j)} \tag{1}$$

respectively for all $\mathbf{s}_i, \mathbf{s}_j \in D$, where $\text{E}[Z(\mathbf{s})]$ denotes the expectation of $Z(\mathbf{s})$. Thus a process is homogeneous if C and R depend only on the separation vector $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$. Furthermore, it is isotropic if C and R depend upon $\mathbf{h}$ only through its length h, i.e.,

$$C(\mathbf{s}_i, \mathbf{s}_j) = C(\mathbf{h}) = C(h), \tag{2}$$

$$R(\mathbf{s}_i, \mathbf{s}_j) = R(\mathbf{h}) = R(h) = C(h)/C(0). \tag{3}$$

If we denote $C(0)$, the variance of $Z(\mathbf{s})$, as σ^2, then the relationship between covariance and correlation is $C(h) = \sigma^2 R(h)$.

Based on the observation in a paper of ICCAD06 [4], the gate length variation in fact has significantly stronger correlation in the x-direction than in the y-direction in reality as in Fig. **??**. Therefore, more accurate anisotropic models are needed.

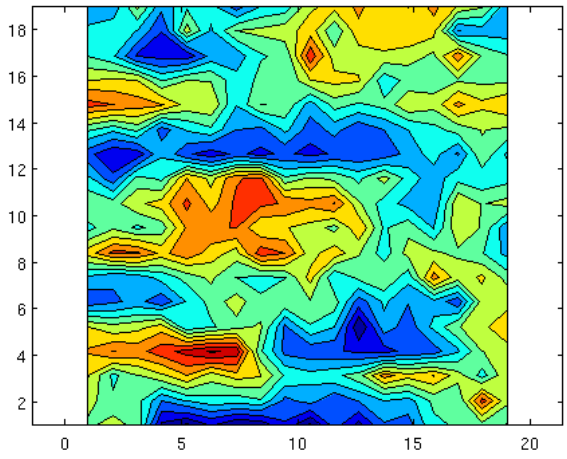

Fig. 1. Four samples of anisotropic data

B. Correlation Function for The Spatially Correlated Component

The spatially correlatied component is modeled as random field with variance σ^2 and correlation function $\rho(h)$ to be extracted. Positive definiteness is the necessary condition for a parametric family of functions to define a legitimate class of correlation functions. Note that the correlation function is an even function, i.e. $\rho(h) = \rho(-h)$, so that its Fourier transform is real. Moreover, because it is positive definite, based on the Bochner's theorem, its Fourier transform is positive. Some process may have further properties:

- monotonicity: correlations are monotonic decreasing against distance.
- nonnegativeness: there is no negative correlaion.

C. Correlation Function Considering Nugget Effect

In the measurement data, apart from the spatially correlated component, the purely random component and the unavoidable measurement error also exist. Typically, the two components are modeled as random variables with independent identical Gaussian distribution, or in short, Gaussian white noise.

When the two components are considered, the measurement data can still be regarded as a Gaussian random field, but the correlation function will have a discontinuity at the origin. This phenomenon is called "nugget effect" [**?**].

III. Non-parametric Estimate

Given N measurement samples $(y_1, y_2, \cdots, y_N) \in \mathbb{R}^n$. The measured covariance matrix $Y == (1/N) \sum_{i=1}^{N} y_i y_i^T$. However, this matrix is unlikely to be positive definite. In [**?**], the following minimization is suggested in order to obtain a nearest positive definite matrix Ω:

$$\begin{array}{ll} \text{minimize} & ||\Omega - Y||_F \\ \text{subject to} & \Omega \succeq 0. \end{array} \tag{4}$$

where $||||_F$ denotes the Frobenius norm and $A \succeq 0$ denotes the matrix A is positive semidefinite. Note that the above problem is convex. In [9], alternate projection method is used. The advantage of alternate projection method is that it is easy to implement. However, only linear convergence can be obtained. More importantly, the method is flexible enough for imposing other constriants. Fortunately, modern optimization techniques can solve the problem effectively and they are publicly available.

Similarly, the maximum likelihood estimation of covariance matrix is given by:

$$\begin{array}{ll} \text{minimize} & \log\det\Omega - \mathrm{Tr}(\Omega^{-1}Y) \\ \text{subject to} & \Omega \succeq 0. \end{array} \tag{5}$$

where $Tr(A)$ denotes the trace of A. Note that the first term is concave whereas second term is convex. However, let $S = \Sigma^{-1}$, the function becomes convex in term of S

$$\begin{array}{ll} \text{maximize} & \log\det S - \mathrm{Tr}(S \cdot Y) \\ \text{subject to} & S \succeq 0, \end{array}$$

However, very often we would like to add extra constraints directly to Ω.We will discuss the MLE estimation in more details in Section **??**. Let us focus on the least squares estimation first.

If the process is known to be isotropic and the correlation model is known in prior, such as Matérn family. The covariance matrix Ω can further be expressed as a non-linear function of parameters, making the problem non-convex. It creates difficulties of numerical solving as mentioned in [?].

In non-parametric approach, however, correlation functions are curved-fitted using linear combination of some basis functions:

$$\rho(h) = \sum_i p_i \Phi_i(h)$$

where p_i's are the unknown coefficients to be fitted and Φ_i's are a family of basis functions. The corresponding covariance matrix Ω can be expressed as an affine function of the coefficients $p_1F_1 + \cdots + p_mF_m$ where $\{F_k\}_{ij} = \Phi_k(s_j - s_i)$. Since it is an affine transformation, convexity is preserved.

Choice of $\Phi_i(h)$:

- $J_i(h)$: Bessel function,
  advantage: positive definite *if and only if* $p_i \geq 0$ for all i.
  disadvantage: difficult to choose i, computational expensive.
- $B_i(h)$: B-spline function,
  advantage: shapes are easier to control,
  e.g. monotonicity ($p_i \geq p_{i+1}$);
  disadvantage: not guarantee positive definite
- $\cos((2\pi i/N)h)$: cosine functions

To ensure that the resulting function is positive definite, additional constraints can be imposed. Based on the Bochner's theorem, we may impose the fast Fourier transform to the discretized function, i.e., each element of

$$\text{real}(\text{FFT}(\{\Phi_i(x_k)\})p$$

is ≥ 0. Note that there is a trade-off of the granularity of discretization about the accuracy of the positive definiteness and efficiency.

IV. B-SPLINE

Let $\mathbf{t} = \{t_j\}_1^{n+k}$ be a nondecreasing sequence called *knot sequence*. The jth (normalized) *B-spline* of order k for $\mathbf{t}$ is denoted by $B_{j,k}$ and is defined by

$$B_{j,k}(x) = (t_{j+k} - t_j)[t_j, \ldots, t_{j+k}](t - x)_+^{k-1}, \quad \text{all } x \in R, \tag{6}$$

which means that $B_{j,k}$ is the multiplication of $(t_{j+k} - t_j)$ and the kth divided difference of the truncated power function $(t - x)_+^{k-1}$.

A spline function $f(x)$ of order k with knot sequence $\mathbf{t}$ is any linear combination of B-splines of order k for the knot sequence $\mathbf{t}$, i.e.,

$$f(x) = \sum_{j=1}^{n} p_j B_{j,k}(x). \tag{7}$$

B-splines have the following fundamental properties [**?**]:

- Non-negative and local support: $B_{j,k}(x)$ is positive if x locates in (t_j, t_{j+k}) and is zero when x is outside this interval;
- Partition of unity: $\sum_{j=1}^{n} B_{j,k}(x) = 1$ if x is in $[t_k, t_{n+1})$;
- Recurrence relation: $B_{j,k}(x)$ can be evaluated recursively by $B_{j,k-1}(x)$ and $B_{j+1,k-1}(x)$.
- Convex hull: For $t_i < x < t_{i+1}$, the value of the spline function $f(x)$ at the site x is a strictly convex combination of the k numbers $p_{i+1-k}, \ldots, p_i$.

This close relationship between the value of a spline and the nearby B-spline coefficients provides the evidence that in order to change part of f's curve we only need to modify its nearby B-spline coefficients. Because of this controllability, in CAGD the term *control point sequence* of the spline function $\sum_j p_j B_{j,k}$ with knot sequence $\mathbf{t}$ is defined as $(t^*_{j,k}, p_j)_{j=1}^n$, where $t^*_{j,k}$ is sometimes called the Greville site or the knot average,

$$t^*_{j,k} = \frac{t_{j+1} + \cdots + t_{j+k+1}}{k+1}, \quad \forall j. \tag{8}$$

For sake of simplicity, we use t^*_j instead as long as the order k is fixed and no confusion is raised hereafter.

If the strictly increasing sequence $\mathbf{x} = (x_i)_1^n$ of data sites is given, then for a given function g, the spline function $f(x)$ defined in (**??**) agrees with g at $\mathbf{x}$ if and only if

$$\sum_{j=1}^{n} p_j B_{j,k}(x_i) = g(x_i), \quad i = 1, \ldots, n. \tag{9}$$

This is a linear system of n equations in the vector $\mathbf{p} = (p_j)_1^n$ of n unknowns, with coefficient matrix $\big(B_{j,k}(x_i)\big)$, the *spline collocation matrix*. Hence the matrix multiplication of (**??**) can be written as $g(\mathbf{x}) = \mathbf{B} \cdot \mathbf{p}$.

By solving out this system we obtain the B-spline function $f(x)$. The advantage of using B-spline for interpolation or approximation lies in the bandedness of collocation matrix with bandwidth less than k. This is really a consequence of

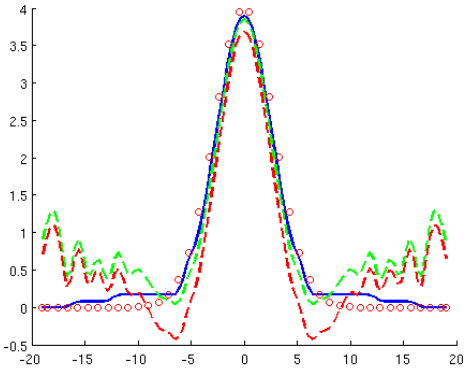

Fig. 2. B-spline fitting with different constraints

Fig. 3. Anisotropic model

B-spline's local support property. Another significant property of the collocation matrix is total positivity, i.e., every non-zero element of $\mathbf{B}$ is positive.

One of the advantages of using B-spline is that the monotonicity can be enforced by imposing a simple linear constraints on p_i, i.e. $p_{i+1} \geq p_i$. The nonnegativeness can be enforced by imposing constraints $p_i \geq 0$. Note however that it is not a neccessary condition so that it may overconstraint the result.

Fig. **??** shows results with different constraints of B-spline fitting.

V. Non-Parametric Estimation for Anisotropic

Result of anisotropic correlation function is in Fig. **??**.

VI. Numerical Experiment

The proposed method was implemented in MATLAB on an Intel® machine with 3.0 GHz XEON™ CPU. Without real silicon measurement data, we synthesized a pseudo measurement process and attained data from a batch of M chips with N sites on each chip. The area of the chip is set to be 10mm$\times$10mm.

VII. Conclusions and Future Directions

Acknowledgment

The authors would like to thank...