Bayesian Model Fusion: Enabling Test Cost Reduction of Analog/RF Circuits via Wafer-level Spatial Variation Modeling

Shanghang Zhang[1], Xin Li[1], R. D. (Shawn) Blanton[1], José Machado da Silva[2],
John M. Carulli[3] and Kenneth M. Butler[3]
[1]ECE Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[2]INESC TEC, Faculdade de Engenharia, Universidade do Porto, 4200-465 Porto, Portugal
[3]Texas Instruments Inc., 12500 TI Boulevard, Dallas, TX 75243, USA
{shanghaz, xinli, blanton}@ece.cmu.edu, jms@fe.up.pt, j.m.carulli@ieee.org, kenb@ti.com

Abstract

In this paper, a novel Bayesian model fusion (BMF) method is proposed for test cost reduction based on wafer-level spatial variation modeling. BMF relies on the assumption that a large number of wafers of the same circuit design (e.g., all wafers from the same lot) share a similar spatial pattern. Hence, the measurement data from one wafer can be borrowed to model the spatial variation of other wafers via Bayesian inference. By applying the Sherman-Morrison-Woodbury formula, a fast numerical algorithm is derived to reduce the computational cost of BMF for practical test applications. Furthermore, a new test methodology is developed based on BMF and it closely monitors the escape rate and yield loss. As is demonstrated by the wafer probe measurement data of an industrial RF transceiver, BMF achieves 1.125× reduction in test cost and 2.6× reduction in yield loss, compared to the conventional approach based on virtual probe (VP).

1. Introduction

The aggressive technology scaling has made integrated circuit (IC) increasingly complex and, hence, difficult to test. Today, a typical system on chip (SOC) is composed of numerous analog/RF and digital circuit blocks. Testing the analog/RF circuits within a complex SOC is extremely expensive and may even dominate the overall test cost [1]-[2]. For this reason, reducing the test cost for analog/RF circuits is one of the grand challenges for the test community.

To address this challenge, a variety of test cost reduction techniques have been proposed during the past one decade [3]-[25]. One of the important ideas is to exploit the *inter-test-item correlation* [3]-[14] so that we do not have to physically measure all test items for a given die. Instead, only a subset of these test items is measured and these measurement results are used to predict other test items that are strongly correlated.

Another alternative approach is to exploit the *spatial correlation* between different dies on the same wafer [15]-[25]. To this end, a number of statistical algorithms and tools, such as virtual probe (VP) [15]-[20] and Gaussian process (GP) [21]-[24], have been developed to model the spatial variation at wafer level. These methods rely on several assumptions that are generally applicable in practice. For instance, VP assumes that the spatial variation of a wafer carries a sparse structure in frequency domain: it can be accurately represented by a small number of spatial frequency components based on discrete cosine transform (DCT) [26]. In this case, we only need to measure a small number of dies from a wafer and then use the measurement results to predict the test items of other dies on the same wafer, thereby significantly reducing the test cost. Furthermore, it has been demonstrated that both the spatial correlation and the inter-test-item correlation can be integrated into a unified statistical framework for test cost reduction [20], [24].

Following this strategy, we propose a novel *Bayesian model fusion* (BMF) technique to efficiently capture the spatial variation at wafer level. Unlike other conventional approaches that rely on the general assumptions of spatial variation (e.g., sparse representation in frequency domain for VP), BMF aims to identify the prior knowledge that is specific to a group of wafers to encode their unique spatial variation information. As such, the proposed BMF technique is expected to model the spatial variation more accurately than other conventional approaches.

In practice, the prior knowledge required by BMF can be extracted by exploiting the fact that a large number of wafers of the same circuit design (e.g., all wafers from the same lot) often share a similar spatial pattern. Hence, we can use one of these wafers to learn our prior knowledge. Once the prior information is available, we only need to measure very few dies on another wafer sharing the similar pattern and combine the measurement data with our prior knowledge via *Bayesian inference* to accurately predict the spatial pattern of that wafer. From this point of view, BMF attempts to model the spatial variation of a wafer by "fusing" the prior knowledge extracted from a previous wafer with the new measurement data collected at the current wafer. Note that we can further incorporate inter-test-item correlation into the proposed BMF framework, even though it is not explicitly studied in this paper and will be considered in our future research.

BMF was initially developed for pre-silicon performance modeling and post-silicon tuning of analog and mixed-signal circuits [27]-[31]. In this paper, we further extend BMF to model spatial variation at wafer level for test cost reduction. In particular, two major contributions are made. First, by following the BMF idea

978-1-4799-4722-5/14/$31.00

in [27]-[31], we derive the mathematical formulation for wafer-level spatial variation modeling. In addition, based upon the *Sherman-Morrison-Woodbury formula* in matrix analysis [32], a fast numerical algorithm is proposed to efficiently solve the Bayesian inference problem posed by BMF for our application of interest.

Second, to facilitate test cost reduction by using BMF, we propose a practical test flow that consists of two phases: (i) *pre-test analysis*, and (ii) *test application*. Pre-test analysis aims to determine whether a given test item is spatially correlated and, hence, the test item is considered to be "predictable". If a test item is predictable, we further determine the number of dies that should be measured on a wafer in order to accurately predict the performance metrics of other unmeasured dies on the same wafer and achieve sufficiently small escape rate and yield loss. Next, at the second phase of test application, we measure the pre-determined number of dies on a wafer for all predictable test items. Meanwhile, we closely monitor the escape rate and yield loss for these predictable test items based on a statistical model that is proposed in this paper. If the escape rate or yield loss exceeds the pre-defined target for a specific wafer (e.g., an outlier wafer with completely different spatial pattern), such an abnormal wafer can be automatically detected by our proposed statistical model and, consequently, all dies on the wafer must be physically measured during test application.

The proposed test flow based on BMF is validated by the wafer probe measurement data of an industrial RF transceiver with 176 wafers and 1,190,816 dies. For this transceiver example, BMF achieves 1.125× reduction in test cost and 2.6× reduction in yield loss over the conventional test flow based on VP.

The remainder of this paper is organized as follows. In Section 2, we derive the mathematical formulation of BMF and then describe the BMF-based test methodology in Section 3. The efficacy of our proposed test cost reduction is demonstrated by the experimental results in Section 4. Finally, we conclude in Section 5.

2. Bayesian Model Fusion

In this section, we first derive the mathematical formulation of the proposed BMF algorithm for spatial variation modeling. Next, three important implementation issues, (i) fast numerical solver, (ii) hyper-parameter estimation, and (iii) reference wafer update, will be further discussed in detail.

2.1 Mathematical Formulation

Consider T test items $\{g_t;\ t = 1, 2, ..., T\}$ (e.g., leakage current, bit error rate, etc.) that should be measured for all dies from M wafers. For each wafer, the measured value of g_t is expected to be different from die to die due to process variation. We use a two-dimensional function $g_{t,m}(x, y)$ to represent the measured value of g_t at the spatial location (x, y) for the m-th wafer where $m \in \{1, 2, ..., M\}$.

Similar to VP [15]-[20], our proposed BMF framework models the wafer-level spatial variation in frequency domain based on DCT:

$$g_{t,m}(x,y) = \sum_{k=1}^{K} \alpha_{t,m,k} \cdot b_k(x,y) + \varepsilon_{t,m}, \tag{1}$$

where $\{b_k(x, y);\ k = 1, 2, ..., K\}$ represents the DCT basis functions, $\{\alpha_{t,m,k};\ k = 1, 2, ..., K\}$ stands for the DCT coefficients, and $\varepsilon_{t,m}$ denotes the modeling error. In practice, the error term $\varepsilon_{t,m}$ models the variation component that is spatially uncorrelated and, hence, cannot be captured by the DCT basis functions in (1).

BMF aims to accurately find the DCT coefficients $\{\alpha_{t,m,k};\ k = 1, 2, ..., K\}$ by measuring very few dies $\{(x^{(n)}, y^{(n)}, g_{t,m}{}^{(n)});\ n = 1, 2, ..., N\}$ from the m-th wafer, where $(x^{(n)}, y^{(n)})$ and $g_{t,m}{}^{(n)}$ denote the spatial location and the measured value of the n-th die respectively and N represents the total number of measured dies. Once these DCT coefficients are known, the spatial variation $g_{t,m}(x, y)$ can be estimated based on the DCT basis functions in (1). We formulate the following linear equations based on the measurement data $\{(x^{(n)}, y^{(n)}, g_{t,m}{}^{(n)});\ n = 1, 2, ..., N\}$:

$$\mathbf{B} \cdot \boldsymbol{\alpha}_{t,m} = \mathbf{g}_{t,m}, \tag{2}$$

where

$$\mathbf{B} = \begin{bmatrix} b_1(x^{(1)},y^{(1)}) & b_2(x^{(1)},y^{(1)}) & \cdots & b_K(x^{(1)},y^{(1)}) \\ b_1(x^{(2)},y^{(2)}) & b_2(x^{(2)},y^{(2)}) & \cdots & b_K(x^{(2)},y^{(2)}) \\ \vdots & \vdots & \vdots & \vdots \\ b_1(x^{(N)},y^{(N)}) & b_2(x^{(N)},y^{(N)}) & \cdots & b_K(x^{(N)},y^{(N)}) \end{bmatrix} \tag{3}$$

$$\boldsymbol{\alpha}_{t,m} = \begin{bmatrix} \alpha_{t,m,1} & \alpha_{t,m,2} & \cdots & \alpha_{t,m,K} \end{bmatrix}^T \tag{4}$$

$$\mathbf{g}_{t,m} = \begin{bmatrix} g_{t,m}^{(1)} & g_{t,m}^{(2)} & \cdots & g_{t,m}^{(N)} \end{bmatrix}^T. \tag{5}$$

Note that the linear equations in (2) are underdetermined, since we often collect very few measurements and, consequently, the number of unknowns (i.e., K) is greater than the number of equations (i.e., N). Because a set of underdetermined linear equations have infinitely many solutions, additional information is required to further constrain the solution space in order to uniquely determine the DCT coefficients $\{\alpha_{t,m,k};\ k = 1, 2, ..., K\}$.

Unlike VP that exploits the sparse structure of $\{\alpha_{t,m,k};\ k = 1, 2, ..., K\}$, our proposed BMF approach relies on a completely different assumption to solve these DCT coefficients. BMF assumes that the M wafers share a similar spatial pattern. Hence, a "model template" can be learned from one of these wafers, referred to as the *reference wafer*, and then applied to all other wafers. In practice, if the spatial patterns of the M wafers are different, we may partition these wafers into multiple sub-groups where all wafers belonging to the same sub-group (e.g., the wafers in the same lot) share a similar pattern. As such, BMF can still be applied to each of these sub-groups individually.

Given a reference wafer (say, the r-th wafer), we first extract its spatial variation $g_{t,r}(x, y)$. Here, the function $g_{t,r}(x, y)$ can be obtained in various ways, e.g., physically

measuring all dies on the wafer, applying the conventional VP method, etc. Once the function $g_{t,r}(x, y)$ is known, we calculate the corresponding DCT coefficients $\{\alpha_{t,r,k};\ k = 1, 2, ..., K\}$. BMF considers these DCT coefficients as the prior knowledge that can be encoded as a prior distribution. Next, for the m-th wafer, BMF combines the prior knowledge with very few measurement data to solve the DCT coefficients $\{\alpha_{t,m,k};\ k = 1, 2, ..., K\}$ by using maximum-a-posterior estimation (MAP) [33]. In what follows, we will describe the details of these two steps, (i) prior definition and (ii) MAP estimation, respectively.

A. Prior Knowledge Definition

Based on the BMF idea in [30], since the reference wafer and the m-th wafer share a similar spatial pattern, we expect that their DCT coefficients $\{\alpha_{t,r,k};\ k = 1, 2, ..., K\}$ and $\{\alpha_{t,m,k};\ k = 1, 2, ..., K\}$ are similar. Such a prior knowledge can be mathematically encoded by the following prior distribution:

$$pdf\left(\alpha_{t,m,k}\right) = \frac{1}{\sqrt{2\pi}\cdot\rho_{t,m}\cdot\left|\alpha_{t,r,k}\right|}\cdot exp\left[-\frac{\left(\alpha_{t,m,k}-\alpha_{t,r,k}\right)^2}{2\cdot\rho_{t,m}^2\cdot\alpha_{t,r,k}^2}\right], \quad (6)$$

where the distribution $pdf(\alpha_{t,m,k})$ is Gaussian, $\alpha_{t,r,k}$ and $\rho_{t,m}^2\cdot\alpha_{t,r,k}^2$ are its mean and variance respectively, and $\rho_{t,m} > 0$ is a hyper-parameter that should be determined.

Studying the prior distribution in (6) reveals several important observations. First, the Gaussian distribution $pdf(\alpha_{t,m,k})$ peaks at its mean value $\alpha_{t,m,k} = \alpha_{t,r,k}$, implying that the DCT coefficients $\alpha_{t,r,k}$ and $\alpha_{t,m,k}$ are likely to be similar. As $\alpha_{t,m,k}$ moves away from its mean value $\alpha_{t,r,k}$, the prior distribution $pdf(\alpha_{t,m,k})$ exponentially decays. Hence, it is unlikely to observe a DCT coefficient $\alpha_{t,m,k}$ that is extremely different from $\alpha_{t,r,k}$. This observation is consistent with our prior knowledge that the spatial patterns of different wafers are similar and, therefore, the corresponding DCT coefficients should be close to each other.

Second, the prior distribution $pdf(\alpha_{t,m,k})$ remains non-zero, even if the DCT coefficients $\alpha_{t,r,k}$ and $\alpha_{t,m,k}$ are not exactly identical. In other words, it is possible for these DCT coefficients to be slightly different. It, in turn, allows us to capture the spatial variation of the m-th wafer that is similar, but not necessarily identical to, the spatial variation of the reference wafer.

Third, the standard deviation of $pdf(\alpha_{t,m,k})$ is proportional to $|\alpha_{t,r,k}|$. Hence, the absolute difference between the DCT coefficients $\alpha_{t,r,k}$ and $\alpha_{t,m,k}$ can be large, if $|\alpha_{t,r,k}|$ is large. Such a definition of prior distribution is adopted, because we want to offer each DCT coefficient $\alpha_{t,m,k}$ a relatively equal opportunity to deviate from $\alpha_{t,r,k}$.

Given the prior distribution of each DCT coefficient $\alpha_{t,m,k}$ in (6), we further assume that all DCT coefficients $\{\alpha_{t,m,k};\ k = 1, 2, ..., K\}$ are statistically independent, resulting in the following joint probability distribution:

$$pdf\left(\boldsymbol{\alpha}_{t,m}\right) = \frac{1}{\left(\sqrt{2\pi}\cdot\rho_{t,m}\right)^K\cdot\prod_{k=1}^{K}\left|\alpha_{t,r,k}\right|}\cdot exp\left[-\frac{\left(\boldsymbol{\alpha}_{t,m}-\boldsymbol{\alpha}_{t,r}\right)^T\cdot\boldsymbol{\Sigma}_t^{-1}\cdot\left(\boldsymbol{\alpha}_{t,m}-\boldsymbol{\alpha}_{t,r}\right)}{2\cdot\rho_{t,m}^2}\right], \quad (7)$$

where $\boldsymbol{\alpha}_{t,r} \in \Re^K$ and $\boldsymbol{\alpha}_{t,m} \in \Re^K$ are defined in (4) and $\boldsymbol{\Sigma}_t \in \Re^{K\times K}$ is a diagonal matrix:

$$\boldsymbol{\Sigma}_t = diag\left(\alpha_{t,r,1}^2 \quad \alpha_{t,r,2}^2 \quad \cdots \quad \alpha_{t,r,K}^2\right). \quad (8)$$

The independence assumption here simply means that we do not know the correlation information as our prior knowledge. However, the mutual correlation between all DCT coefficients $\{\alpha_{t,m,k};\ k = 1, 2, ..., K\}$ will be taken into account, once we combine the prior distribution $pdf(\boldsymbol{\alpha}_{t,m})$ in (7) with the measurement data $\{(x^{(n)}, y^{(n)}, g_{t,m}{}^{(n)});\ n = 1, 2, ..., N\}$ for MAP estimation. The details of MAP will be discussed in the next sub-section.

B. Maximum-a-posterior Estimation

The prior distribution $pdf(\boldsymbol{\alpha}_{t,m})$ in (7) tells us the possible values that the DCT coefficients $\{\alpha_{t,m,k};\ k = 1, 2, ..., K\}$ may take. Namely, the DCT coefficients are likely to take the values at which $pdf(\boldsymbol{\alpha}_{t,m})$ is large. On the other hand, since $pdf(\boldsymbol{\alpha}_{t,m})$ defines a distribution that may spread over a wide range, we cannot uniquely determine the DCT coefficients $\{\alpha_{t,m,k};\ k = 1, 2, ..., K\}$ based on the prior distribution $pdf(\boldsymbol{\alpha}_{t,m})$ only. Additional information must be provided by collecting the measurement data from the m-th wafer in order to solve these unknown DCT coefficients.

Assume that N dies are physically measured from the m-th wafer, resulting in the data set $\{(x^{(n)}, y^{(n)}, g_{t,m}{}^{(n)});\ n = 1, 2, ..., N\}$. The measurement data can further be combined with our prior knowledge to reduce the uncertainties associated with the DCT coefficients $\{\alpha_{t,m,k};\ k = 1, 2, ..., K\}$. Based on Bayes' theorem [33], the uncertainties of $\{\alpha_{t,m,k};\ k = 1, 2, ..., K\}$ after knowing $\{(x^{(n)}, y^{(n)}, g_{t,m}{}^{(n)});\ n = 1, 2, ..., N\}$ can be mathematically described by the following posterior distribution:

$$pdf\left(\boldsymbol{\alpha}_{t,m}\middle|\mathbf{g}_{t,m}\right) \propto pdf\left(\boldsymbol{\alpha}_{t,m}\right)\cdot pdf\left(\mathbf{g}_{t,m}\middle|\boldsymbol{\alpha}_{t,m}\right), \quad (9)$$

where $\boldsymbol{\alpha}_{t,m} \in \Re^K$ and $\mathbf{g}_{t,m} \in \Re^N$ are defined in (4) and (5) respectively. In (9), the prior distribution $pdf(\boldsymbol{\alpha}_{t,m})$ is specified by (7), and the conditional distribution $pdf(\mathbf{g}_{t,m} \mid \boldsymbol{\alpha}_{t,m})$ stands for the likelihood function that specifies the probability of observing the measurement data $\{(x^{(n)}, y^{(n)}, g_{t,m}{}^{(n)});\ n = 1, 2, ..., N\}$ for a set of given DCT coefficients $\{\alpha_{t,m,k};\ k = 1, 2, ..., K\}$.

To derive the mathematical representation of the likelihood function $pdf(\mathbf{g}_{t,m} \mid \boldsymbol{\alpha}_{t,m})$, we re-write the DCT equation in (1) as:

$$\varepsilon_{t,m} = g_{t,m}\left(x,y\right) - \sum_{k=1}^{K}\alpha_{t,m,k}\cdot b_k\left(x,y\right). \quad (10)$$

We statistically model the error term $\varepsilon_{t,m}$ in (10) as a zero-mean Gaussian distribution:

$$pdf\left(\varepsilon_{t,m}\right)=\frac{1}{\sqrt{2\pi}\cdot\sigma_{t,m}}\cdot exp\left[-\frac{\varepsilon_{t,m}^2}{2\cdot\sigma_{t,m}^2}\right], \quad (11)$$

where the standard deviation $\sigma_{t,m}$ is another hyper-parameter that should be determined. Hence, the likelihood of observing the measured value $g_{t,m}^{(n)}$ for a particular die at the spatial location $(x^{(n)}, y^{(n)})$ is equal to:

$$pdf\left[g_{t,m}^{(n)}\middle|\boldsymbol{\alpha}_{t,m}\right]=\frac{1}{\sqrt{2\pi}\cdot\sigma_{t,m}}\cdot exp\left\{-\frac{1}{2\cdot\sigma_{t,m}^2}\cdot\left[g_{t,m}^{(n)}-\sum_{k=1}^{K}\alpha_{t,m,k}\cdot b_k\left(x^{(n)},y^{(n)}\right)\right]^2\right\} \quad (12)$$

In addition, since the error term $\varepsilon_{t,m}$ in (10) models the uncorrelated spatial variation in our application, its values should be statistically independent at different spatial locations. For this reason, the likelihood of observing all measurement data $\{(x^{(n)}, y^{(n)}, g_{t,m}^{(n)}); n = 1, 2, ..., N\}$ can be expressed as:

$$pdf\left(\mathbf{g}_{t,m}\middle|\boldsymbol{\alpha}_{t,m}\right)=\prod_{n=1}^{N}pdf\left[g_{t,m}^{(n)}\middle|\boldsymbol{\alpha}_{t,m}\right]=\frac{1}{\left(\sqrt{2\pi}\cdot\sigma_{t,m}\right)^N}\cdot exp\left\{-\frac{\left(\mathbf{g}_{t,m}-\mathbf{B}\cdot\boldsymbol{\alpha}_{t,m}\right)^T\cdot\left(\mathbf{g}_{t,m}-\mathbf{B}\cdot\boldsymbol{\alpha}_{t,m}\right)}{2\cdot\sigma_{t,m}^2}\right\}, \quad (13)$$

where $\mathbf{B} \in \Re^{N\times K}$ is defined in (3).

Substituting (7) and (13) into (9) yields the posterior distribution:

$$pdf\left(\boldsymbol{\alpha}_{t,m}\middle|\mathbf{g}_{t,m}\right)\propto exp\left[-\frac{\left(\boldsymbol{\alpha}_{t,m}-\boldsymbol{\alpha}_{t,r}\right)^T\cdot\boldsymbol{\Sigma}_t^{-1}\cdot\left(\boldsymbol{\alpha}_{t,m}-\boldsymbol{\alpha}_{t,r}\right)}{2\cdot\rho_{t,m}^2}-\frac{\left(\mathbf{g}_{t,m}-\mathbf{B}\cdot\boldsymbol{\alpha}_{t,m}\right)^T\cdot\left(\mathbf{g}_{t,m}-\mathbf{B}\cdot\boldsymbol{\alpha}_{t,m}\right)}{2\cdot\sigma_{t,m}^2}\right] \quad (14)$$

The posterior distribution $pdf(\boldsymbol{\alpha}_{t,m} \mid \mathbf{g}_{t,m})$ in (14) models the uncertainties of the DCT coefficients $\{\alpha_{t,m,k}; k = 1, 2, ..., K\}$, after we observe the measurement data $\{(x^{(n)}, y^{(n)}, g_{t,m}^{(n)}); n = 1, 2, ..., N\}$. Namely, even with the measurement data, the DCT coefficients are still not deterministic. However, based on $pdf(\boldsymbol{\alpha}_{t,m} \mid \mathbf{g}_{t,m})$ in (14), we can find the values of the DCT coefficients that are most likely to occur, and consider these values as the optimal estimation of $\{\alpha_{t,m,k}; k = 1, 2, ..., K\}$. Such an approach is referred to as the MAP estimation in the statistics community [33].

Following this idea, we need to solve the following optimization problem:

$$\underset{\boldsymbol{\alpha}_{t,m}}{\text{maximize}} \quad pdf\left(\boldsymbol{\alpha}_{t,m}\middle|\mathbf{g}_{t,m}\right), \quad (15)$$

where the posterior distribution $pdf(\boldsymbol{\alpha}_{t,m} \mid \mathbf{g}_{t,m})$ in (14) is the merit function to be maximized. Based on the first-order optimality condition, the gradient of the merit function with respect to $\boldsymbol{\alpha}_{t,m}$ should be zero at the optimal solution [34]. Hence, it is straightforward to derive the following solution:

$$\boldsymbol{\alpha}_{t,m}=\left(\eta_{t,m}\cdot\boldsymbol{\Sigma}_t^{-1}+\mathbf{B}^T\cdot\mathbf{B}\right)^{-1}\cdot\left(\eta_{t,m}\cdot\boldsymbol{\Sigma}_t^{-1}\cdot\boldsymbol{\alpha}_{t,r}+\mathbf{B}^T\cdot\mathbf{g}_{t,m}\right), \quad (16)$$

where

$$\eta_{t,m}=\sigma_{t,m}^2/\rho_{t,m}^2. \quad (17)$$

Studying (17) reveals an important fact that even though there are two hyper-parameters (i.e., $\rho_{t,m}$ and $\sigma_{t,m}$) in our Bayesian formulation, we only need to know their ratio $\eta_{t,m} = \sigma_{t,m}^2 / \rho_{t,m}^2$ in order to solve the DCT coefficients $\{\alpha_{t,m,k}; k = 1, 2, ..., K\}$ by MAP. Once these DCT coefficients are determined, the spatial variation of the m-th wafer can be estimated by (1). In our implementation, the hyper-parameter $\eta_{t,m}$ is found by using the cross-validation method, as will be discussed in Section 2.3.

2.2 Fast Numerical Solver

As shown in (16), we need to solve a set of linear equations to determine the unknown DCT coefficients $\{\alpha_{t,m,k}; k = 1, 2, ..., K\}$. In practice, if a wafer contains many dies and, consequently, many DCT basis functions must be used to model the wafer-level spatial variation with high resolution, there are a large number of DCT coefficients that we have to solve from (16). For example, as will be demonstrated by an industrial RF transceiver in Section 4, there are more than 6700 dies on each wafer, since the silicon area occupied by a transceiver circuit is small. In this case, we need to solve thousands of the DCT coefficients from (16), resulting in expensive computational time. Note that computational cost is a critical issue for our application, since we must run the proposed BMF algorithm in real time during the testing process.

To address this technical challenge of complexity, we propose a novel numerical algorithm to substantially reduce the computational time of solving (16). Studying (16), we have two important observations. First, the matrix $\boldsymbol{\Sigma}_t \in \Re^{K\times K}$ is diagonal and its inverse can be easily calculated with low computational cost. Second, the matrix $\mathbf{B} \in \Re^{N\times K}$ has more columns than rows, since the number of measured dies (i.e., N) is less than the number of unknown DCT coefficients (i.e., K) in our application. Hence, the matrix $\mathbf{B}^T\cdot\mathbf{B} \in \Re^{K\times K}$ is not full-rank. Its rank N is much less than its size K. We can adopt the Sherman-Morrison-Woodbury formula [32] to efficiently calculate the following inverse matrix based on low-rank update:

$$\left(\eta_{t,m}\cdot\boldsymbol{\Sigma}_t^{-1}+\mathbf{B}^T\cdot\mathbf{B}\right)^{-1}=\eta_{t,m}^{-1}\boldsymbol{\Sigma}_t-\eta_{t,m}^{-1}\boldsymbol{\Sigma}_t\mathbf{B}^T\cdot\left(\eta_{t,m}\mathbf{I}+\mathbf{B}\boldsymbol{\Sigma}_t\mathbf{B}^T\right)^{-1}\cdot\mathbf{B}\boldsymbol{\Sigma}_t, \quad (18)$$

where $\mathbf{I} \in \Re^{N\times N}$ denotes an identity matrix. Substituting (18) into (16) yields:

$$\boldsymbol{\alpha}_{t,m}=\eta_{t,m}^{-1}\boldsymbol{\Sigma}_t\cdot\left(\eta_{t,m}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\alpha}_{t,r}+\mathbf{B}^T\mathbf{g}_{t,m}\right)-\eta_{t,m}^{-1}\boldsymbol{\Sigma}_t^{-1}\mathbf{B}^T\cdot\left(\eta_{t,m}\mathbf{I}+\mathbf{B}\boldsymbol{\Sigma}_t\mathbf{B}^T\right)^{-1}\cdot\mathbf{B}\boldsymbol{\Sigma}_t\cdot\left(\eta_{t,m}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\alpha}_{t,r}+\mathbf{B}^T\mathbf{g}_{t,m}\right). \quad (19)$$

Note that the expression in (19) does not rely on any approximation. Hence, it results in the exact solution of $\boldsymbol{\alpha}_{t,m}$ except for numerical errors.

Algorithm 1: Fast Numerical Solver for BMF

1. Start from $\boldsymbol{\alpha}_{t,r} \in \Re^K$, $\mathbf{g}_{t,m} \in \Re^N$, $\mathbf{B} \in \Re^{N\times K}$, $\boldsymbol{\Sigma}_t \in \Re^{K\times K}$ and a given value of $\eta_{t,m}$.
2. Calculate $\mathbf{p} = \eta_{t,\mathrm{m}}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\alpha}_{t,r} + \mathbf{B}^T\mathbf{g}_{t,m}$, where $\boldsymbol{\Sigma}_t$ is a diagonal matrix.
3. Calculate $\mathbf{q} = (\eta_{t,\mathrm{m}}\mathbf{I} + \mathbf{B}\boldsymbol{\Sigma}_t\mathbf{B}^T)^{-1}\cdot\mathbf{B}\boldsymbol{\Sigma}_t\mathbf{p}$ by solving N linear equations.
4. Calculate $\boldsymbol{\alpha}_{t,m} = \eta_{t,m}^{-1}\boldsymbol{\Sigma}_t\mathbf{p} - \eta_{t,\mathrm{m}}^{-1}\boldsymbol{\Sigma}_t\mathbf{B}^T\mathbf{q}$.

Algorithm 1 summarizes the major steps of the proposed fast solver. Even though Algorithm 1 involves more steps than directly solving (16) by Gaussian elimination, it is more computationally efficient than the conventional solver because all steps in Algorithm 1 are computationally inexpensive. Most importantly, unlike (16) that must solve K linear equations, Algorithm 1 only solves N linear equations in Step 3 where N is much less than K in our application. It, in turn, explains the reason why Algorithm 1 is preferred over the conventional solver. As will be demonstrated by our experimental results in Section 4, the proposed fast solver achieves up to 166× runtime speed-up over the conventional solver based on Gaussian elimination.

2.3 Hyper-parameter Estimation

As shown in (16), the DCT coefficients $\{\alpha_{t,m,k}; k = 1, 2, ..., K\}$ estimated by MAP depends on the hyper-parameter $\eta_{t,m}$. Note that if $\eta_{t,m}$ is too small (i.e., $\eta_{t,m} \to 0$), the solution $\boldsymbol{\alpha}_{t,m}$ is almost independent of $\boldsymbol{\alpha}_{t,r}$ and $\boldsymbol{\Sigma}_t$, that encode the prior knowledge. Namely, the prior information is neglected. Hence, the resulting DCT coefficients are likely to be inaccurate due to over-fitting, because we have more unknown coefficients than measured dies and the linear equations in (2) are underdetermined. On the other hand, if $\eta_{t,m}$ is too large (i.e., $\eta_{t,m} \to +\infty$), the solution $\boldsymbol{\alpha}_{t,m}$ simply equals $\boldsymbol{\alpha}_{t,r}$. It, in turn, implies that the difference between the reference wafer and the m-th wafer is not captured and, therefore, the solution $\boldsymbol{\alpha}_{t,m}$ is not accurate either.

The aforementioned discussion indicates that we must appropriately determine the hyper-parameter $\eta_{t,m}$ in order to accurately solve the DCT coefficients by (16). Ideally, we want to find the optimal $\eta_{t,m}$ that results in the minimal modeling error. However, it is not trivial to estimate the modeling error based on the measurement data $\{(x^{(n)}, y^{(n)}, g_{t,m}{}^{(n)}); n = 1, 2, ..., N\}$. Remember that we only measure very few dies (i.e., N is extremely small) and the linear equations in (2) are underdetermined. In this case, even if the fitted DCT coefficients perfectly satisfy the linear equations in (2), our spatial variation model may over-fit the measurement data and it may not accurately predict the performance metrics of other unmeasured dies. For this reason, we must adopt an error estimation strategy that does not suffer from the over-fitting problem.

Towards this goal, we borrow the cross-validation approach from the statistics community [33]. The key idea is not to estimate the modeling error based on the same data set that is used to solve the DCT coefficients. Instead, the modeling error must be estimated from an independent data set. For this reason, an S-fold cross-validation partitions the measurement data $\{(x^{(n)}, y^{(n)}, g_{t,m}{}^{(n)}); n = 1, 2, ..., N\}$ into S non-overlapped groups, and repeatedly run the proposed BMF algorithm for S times. At each run, one of these S groups is used to estimate the modeling error, while all other groups are used to solve the DCT coefficients. As such, since the spatial variation model is fitted and then validated by non-overlapped data sets, over-fitting can be reliably detected. The overall modeling error is calculated by averaging the individual modeling error over these S runs. More details of cross-validation can be found in [33].

By applying the cross-validation method, we can estimate the modeling error associated with different values of $\eta_{t,m}$. Next, we choose the optimal $\eta_{t,m}$ at which the modeling error reaches its minimum. The aforementioned cross-validation approach requires to repeatedly run the BMF algorithm for multiple times. However, since each run is computationally inexpensive by taking advantage of the fast solver proposed in Section 2.2, the overall computational cost remains tractable, as will be demonstrated by our experimental example in Section 4.

2.4 Reference Wafer Update

Our proposed BMF method requires the extraction of prior knowledge from a reference wafer. In practice, it is not necessary to use a fixed reference wafer to predict the spatial variation of all other wafers. Instead, we can dynamically update the choice of reference wafer to minimize the difference between the reference wafer and the wafer that we aim to predict the spatial variation. As such, the modeling accuracy of BMF can be maximized.

In this paper, we assume that all wafers are sorted by their temporal sequence and/or spatial location. For instance, if these wafers come from the same lot, they can be sorted by their spatial location in the lot. As a result, two adjacent wafers are likely to share a similar spatial pattern, since they almost sit at the same physical location of the same lot.

When applying the proposed BMF method, we initially measure all dies on the first wafer to obtain its spatial variation $g_{t,1}(x, y)$. Next, we consider the first wafer as the reference wafer to predict the spatial variation of the second wafer $g_{t,2}(x, y)$. Once $g_{t,2}(x, y)$ is known, we further consider the second wafer as the reference wafer to predict the spatial variation of the third wafer $g_{t,3}(x, y)$. Following this iterative scheme, after we know the spatial variation of the m-th wafer $g_{t,m}(x, y)$, we consider the m-th wafer as our reference wafer to predict the spatial variation of the $(m+1)$-th wafer $g_{t,m+1}(x, y)$. By dynamically updating the choice of reference wafer, we can now track the process shift and/or other systematic wafer-to-wafer variation by the proposed BMF framework.

3. Proposed Test Methodology

Our proposed test methodology based on BMF consists

of two phases: (i) pre-test analysis, and (ii) test application. In this section, we will describe the details of these two phases and highlight their novelty.

3.1 Pre-test Analysis

The objective of pre-test analysis is to determine whether a test item g_t is spatially correlated and, hence, can be predicted by BMF. If the test item g_t is predictable, we will further determine the number of dies that should be physically measured in order to accurately model its spatial variation. Otherwise, if the test item g_t is unpredictable, we must physically measure all dies on the wafer to determine whether g_t passes or fails the performance specification for each die.

Given M wafers in total, we measure all dies from the first wafer and use the measurement data for pre-test analysis. The proposed pre-test analysis is composed of two critical components: (i) estimating escape rate and yield loss, and (ii) assessing predictability.

A. Estimating Escape Rate and Yield Loss

The required accuracy of spatial variation modeling is closely related to the target of escape rate and yield loss. For example, if a test item g_t is assigned with a tight performance specification and it is the limiting factor of the overall parametric yield, we must accurately model its spatial variation so that the escape rate and yield loss are sufficiently small. It, in turn, implies that we must assess the predictability of a test item g_t based on its impact on escape rate and yield loss. Therefore, accurately estimating the escape rate and yield loss for each test item is an important task for our proposed pre-test analysis.

Without loss of generality, we assume that each test item g_t, where $t \in \{1, 2, ..., T\}$, is associated with a lower bound l_t and an upper bound u_t. The test item g_t of a die must be within the interval $[l_t, u_t]$ so that the die is considered to pass the performance specification. The escape rate ER_t and the yield loss YL_t associated with the t-th test item are defined as:

$$ER_t = prob\left[\left(g_t \le l_t \text{ or } g_t \ge u_t\right) \quad and \quad l_t \le \tilde{g}_t \le u_t\right] \quad (20)$$

$$YL_t = prob\left[l_t \le g_t \le u_t \quad and \quad \left(\tilde{g}_t \le l_t \text{ or } \tilde{g}_t \ge u_t\right)\right], \quad (21)$$

where g_t and $\tilde{g}_t$ denote the actual and predicted values of the t-th test item respectively, and $prob(\bullet)$ stands for the probability of a random event. Note that the escape rate and yield loss in (20)-(21) are defined for a given test item g_t. It is not the overall escape rate and yield loss at the die level.

When BMF is applied, a number of (say, N) dies are measured from the m-th wafer for the t-th test item. This data set $\{(x^{(n)}, y^{(n)}, g_{t,m}{}^{(n)}); n = 1, 2, ..., N\}$ is used to predict the spatial variation $g_{t,m}(x, y)$. Based on the cross-validation method described in Section 2.3, we obtain a predicted value $\tilde{g}_{t,m}$ at each measurement location $\{(x^{(n)}, y^{(n)})$ where $n \in \{1, 2, ..., N\}$. In other words, we now have N data points $\{(g_{t,m}{}^{(n)}, \tilde{g}_{t,m}{}^{(n)}); n = 1, 2, ..., N\}$ for the actual and predicted values of the t-th test item.

To estimate the escape rate and yield loss, we model $g_{t,m}$ and $\tilde{g}_{t,m}$ as a bivariate normal distribution:

$$pdf\left(g_{t,m}, \tilde{g}_{t,m}\right) = \frac{1}{2\pi \cdot \sqrt{det\left(\mathbf{\Sigma}_{t,m}\right)}} \cdot exp\left\{-\frac{1}{2} \cdot \left(\begin{bmatrix} g_{t,m} \\ \tilde{g}_{t,m} \end{bmatrix} - \mathbf{\mu}_{t,m}\right)^T \cdot \mathbf{\Sigma}_{t,m}^{-1} \cdot \left(\begin{bmatrix} g_{t,m} \\ \tilde{g}_{t,m} \end{bmatrix} - \mathbf{\mu}_{t,m}\right)\right\}, \quad (22)$$

where $det(\bullet)$ denotes the determinant of a matrix, and the mean vector $\mathbf{\mu}_{t,m}$ and the covariance matrix $\mathbf{\Sigma}_{t,m}$ are estimated by the N data points $\{(g_{t,m}{}^{(n)}, \tilde{g}_{t,m}{}^{(n)}); n = 1, 2, ..., N\}$:

$$\mathbf{\mu}_{t,m} = \frac{1}{N} \cdot \sum_{n=1}^{N} \begin{bmatrix} g_{t,m}^{(n)} \\ \tilde{g}_{t,m}^{(n)} \end{bmatrix} \quad (23)$$

$$\mathbf{\Sigma}_{t,m} = \frac{1}{N-1} \cdot \sum_{n=1}^{N} \left(\begin{bmatrix} g_{t,m}^{(n)} \\ \tilde{g}_{t,m}^{(n)} \end{bmatrix} - \mathbf{\mu}_{t,m}\right) \cdot \left(\begin{bmatrix} g_{t,m}^{(n)} \\ \tilde{g}_{t,m}^{(n)} \end{bmatrix} - \mathbf{\mu}_{t,m}\right)^T. \quad (24)$$

Once the distribution in (22) is known, the escape rate $ER_{t,m}$ and the yield loss $YL_{t,m}$ of the t-th test item for the m-th wafer can be calculated by the following two-dimensional numerical integration [35]:

$$ER_{t,m} = \int\limits_{\substack{g_{t,m} \le l_t \text{ or } g_{t,m} \ge u_t \\ l_t \le \tilde{g}_{t,m} \le u_t}} pdf\left(g_{t,m}, \tilde{g}_{t,m}\right) \cdot dg_{t,m} \cdot d\tilde{g}_{t,m} \quad (25)$$

$$YL_{t,m} = \int\limits_{\substack{l_t \le g_{t,m} \le u_t \\ \tilde{g}_{t,m} \le l_t \text{ or } \tilde{g}_{t,m} \ge u_t}} pdf\left(g_{t,m}, \tilde{g}_{t,m}\right) \cdot dg_{t,m} \cdot d\tilde{g}_{t,m}. \quad (26)$$

The aforementioned approach for estimating escape rate and yield loss is based upon normal distribution. In practice, a test item g_t may not be normally distributed. In this case, as long as we know its probability distribution, we can apply the same idea to estimate the escape rate and yield loss. Namely, we first use the N data points $\{(g_{t,m}{}^{(n)}, \tilde{g}_{t,m}{}^{(n)}); n = 1, 2, ..., N\}$ to estimate the important parameters (e.g., mean, co-variance, etc.) of the distribution and then apply numerical integration to calculate the escape rate and yield loss.

B. Assessing Predictability

Once we know how to estimate the escape rate and yield loss for BMF, we can assess the predictability for each test item g_t where $t \in \{1, 2, ..., T\}$. The test item g_t is predictable, if and only if its escape rate and yield loss are both sufficiently small (i.e., less than the pre-defined target) when N dies are physically measured only.

Note that our definition of predictability depends on the number of measured dies. If very few dies are measured for the test item g_t, the spatial variation $g_{t,m}(x, y)$ cannot be accurately predicted and, hence, the test item g_t is considered to be unpredictable. On the other hand, as the number of measured dies increases and, consequently, the modeling accuracy of $g_{t,m}(x, y)$ is improved, the test item g_t may become predictable. For this reason, we must first appropriately choose the number of measured dies (i.e., N), before the predictability of a test item g_t is concluded.

For production test in practice, it is expensive, if not

impossible, to measure a completely different set of test items for different dies on the same wafer. We only have limited flexibility to customize the test scheme for different dies at different spatial locations. In other words, it is not practically feasible to choose many different values of N for different test items. Instead, a large number of test items must share the same value of N so that they can be measured together from the same set of N dies.

With this constraint, we propose to adopt a simple strategy where we apply a fixed value of N to all test items $\{g_t;\ t = 1, 2, ..., T\}$ and then classify these test items into two different categories: predictable and unpredictable. During production test, all test items will be measured from N dies over a wafer. Next, the unpredictable test items will be further measured from the other dies on the same wafer. Such a test scheme can be easily implemented with low cost.

In order to find the optimal value of N that minimizes the test cost, we perform linear search over all possible values of N. In practice, the test cost can be measured by the corresponding test time. In this paper, since the test time for each test item is not disclosed by our industrial collaborator for the experimental example in Section 4, we simply use the total number of measurements to assess the test cost. For instance, if we need to measure T test items from N dies, the test cost is counted as $T{\cdot}N$.

The proposed linear search method estimates the test cost for each possible value of N based on the measurement data collected from the first wafer. For a given value of N, we apply BMF to the first wafer where the same wafer is also used as the reference wafer. If a test item g_t is predictable with sufficiently small escape rate and yield loss, only N dies should be physically measured for g_t on a wafer. Otherwise, if a test item g_t is considered to be unpredictable, all dies on the wafer must be physically measured. We count the total number of measurements for all test items of all dies over the first wafer and use it as a metric to quantitatively measure the test cost. The aforementioned procedure is repeated for different values of N. Finally, we choose the optimal value of N at which the test cost reaches the minimum.

3.2 Test Application

Once the pre-test analysis is complete, each test item g_t is labelled as either predictable or unpredictable. During the second phase of test application, we first physically measure all test items from N dies of a wafer (say, the m-th wafer). For each predictable test item g_t, we apply BMF to predict the spatial variation $g_{t,m}(x, y)$ based on the measurement data collected from these N dies. Next, we check the escape rate and yield loss by applying the statistical model proposed in Section 3.1.A. If the escape rate or yield loss exceeds the pre-defined target, it implies that the spatial variation $g_{t,m}(x, y)$ of the m-th wafer may carry a different pattern and cannot be accurately predicted by BMF with N dies. In this case, the test item g_t is temporarily labelled as unpredictable for the m-th wafer.

Finally, we measure the unpredictable test items from all other dies on the m-th wafer. Here, a test item g_t is considered to be unpredictable, if it is classified as an unpredictable test item during the pre-test analysis or its escape rate or yield loss exceeds the pre-defined target for the m-th wafer. In either case, all dies on the m-th wafer are physically measured for g_t. The aforementioned test scheme is repeated until all wafers are tested.

3.3 Summary

Algorithm 2 summarizes the major steps of the proposed flow for test cost reduction based on BMF. Starting from M wafers that share a similar spatial pattern, we first physically measure all dies from the first wafer and use the measurement data for pre-test analysis. Next, we test all other wafers one by one, where only N dies are physically measured for the predictable test items $\{g_t;\ t \in \Omega\}$ on each wafer. Meanwhile, we closely monitor the escape rate and yield loss associated with these predictable test items. If the estimated escape rate or yield loss exceeds the pre-defined target for the predictable test item g_t of a given wafer, all dies on that wafer will be physically measured for g_t.

Algorithm 2: Test Cost Reduction by BMF

1. Start from M wafers that share a similar spatial pattern.
2. Physically measure all dies from the first wafer. Perform pre-test analysis to determine the set of predictable test items $\{g_t;\ t \in \Omega\}$ and the optimal number of measured dies (i.e., N).
3. For $m = 2, 3, ..., M$
4. Initialize the set $\Theta = \{\}$.
5. Physically measure all test items from N randomly selected dies on the m-th wafer.
6. For each test item g_t where $t \in \Omega$, consider the $(m-1)$-th wafer as the reference wafer and apply BMF to predict the spatial variation $g_{t,m}(x, y)$ of the m-th wafer. Estimate the escape rate $ER_{t,m}$ and the yield loss $YL_{t,m}$. If $ER_{t,m}$ or $YL_{t,m}$ exceeds the pre-defined target, set $\Theta = \Theta \cup \{t\}$.
7. Physically measure the test items $\{g_t;\ t \notin \Omega$ or $t \in \Theta\}$ for all other dies on the m-th wafer.
8. Determine "pass" or "fail" for each die on the m-th wafer.
9. End For

It is important to mention that if the escape rate or yield loss of a test item g_t becomes extremely large for a large number of wafers, it often implies that the spatial variation associated with g_t has been significantly changed. In this case, we should redo the pre-test analysis in order to accommodate the new spatial variation. In addition, even though Algorithm 2 only exploits the spatial correlation for test cost reduction, the proposed BMF framework can be extended to incorporate the additional information of inter-test-item correlation. The details of these ideas are not studied in this paper and will be considered in our future research.

4. Experimental Results

In this section, we demonstrate the efficacy of the proposed test cost reduction scheme by using the wafer probe measurement data of an industrial RF transceiver with 1,190,816 dies. These dies are distributed over 176 wafers and 9 lots where each wafer contains 6,766 dies. For each die, 51 test items (e.g., bit error rate, power consumption, standby current, etc.) are considered in our experiment.

For comparison purposes, two different schemes are implemented for test cost reduction based on (i) VP and (ii) BMF, respectively. Both schemes follow the same test flow described by Algorithm 2 and only their approaches for spatial variation modeling are different. All numerical experiments are performed on a Linux server with 3.4 GHz CPU and 16 GB memory.

4.1 Bayesian Model Fusion

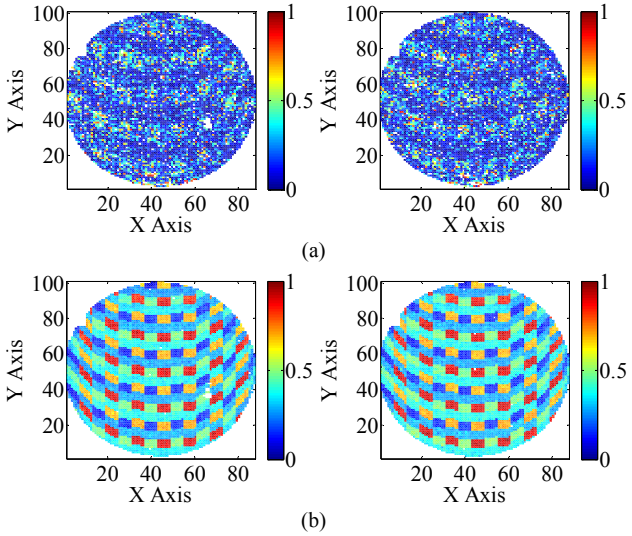

Figure 1. Normalized spatial variation is shown for (a) two wafers of test item #1 that is spatially uncorrelated, and (b) two wafers of test item #48 that is spatially correlated.

Figure 1 shows several examples of wafer-level spatial variation. Based on these plots, two important observations can be made. First, not all test items are spatially correlated. As shown in Figure 1(a), the test item #1 does not carry a clear spatial pattern and, hence, its test cost cannot be reduced by exploiting the spatial correlation information. Second, but more importantly, a number of other test items are indeed spatially correlated. In addition, the spatial patterns of these test items are similar across multiple wafers, as shown in Figure 1(b). It, in turn, validates the fundamental assumption of the proposed BMF method.

To demonstrate the superior modeling accuracy of BMF, we measure all test items from 100 dies on a wafer. Next, we apply both VP and BMF to predict the spatial variation of each test item with 8,888 DCT basis functions. Figure 2(a) compares the modeling error between VP and BMF where the VP error is considered as the baseline. The modeling error of the t-th test item for the m-th wafer is defined as:

$$Error_{t,m} = \sqrt{\sum_{n} \left(g_{t,m}^{(n)} - \tilde{g}_{t,m}^{(n)} \right)^2}, \quad (27)$$

where $g_{t,m}^{(n)}$ and $\tilde{g}_{t,m}^{(n)}$ denote the actual and predicted values of the test item for the n-th die respectively, and the summation in (27) is calculated over all dies on the wafer. Note that BMF achieves up to 300% error reduction in this example. Figure 2(b)-(d) further show the actual, VP-predicted and BMF-predicted spatial variation for the test item #48. It is straightforward to observe that BMF accurately captures the spatial variation with 100 measured dies only, while VP completely fails to work in this example.

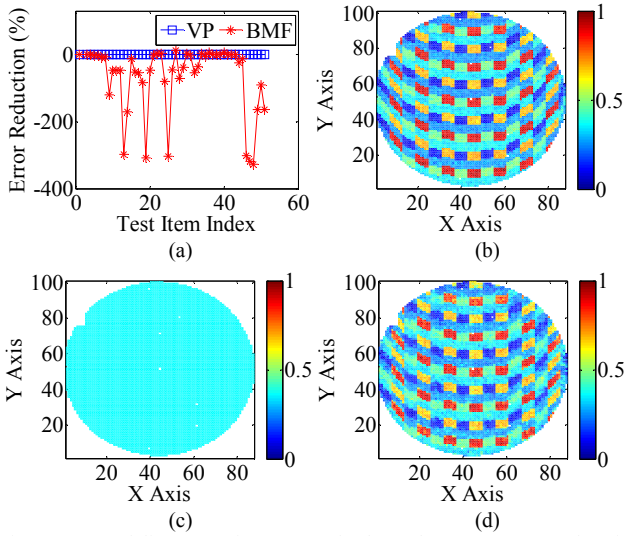

Figure 2. Modeling error is compared where the measurement data is collected from 100 dies on a wafer. (a) Modeling error is compared between VP and BMF for all test items where the VP error is considered as the baseline. (b)-(d) Normalized actual (b), VP-predicted (c) and BMF-predicted (d) spatial variation is shown for the test item #48.

Table 1. Computational time of BMF with cross-validation to model the spatial variation of a single test item for a given wafer

Number of Measured Dies (N)	Computational Time (Sec.)	
	Direct Solver (Conventional)	Fast Solver (Proposed)
100	216.6	1.3
250	222.5	3.3
500	234.8	7.4
1000	257.1	18.3
2000	302.3	54.3
4000	396.0	180.6

Table 1 compares the computational time between the conventional direct solver based on Gaussian elimination and the proposed fast solver based on the Sherman-Morrison-Woodbury formula. The computational time is shown for different numbers of measured dies (i.e., N). As the value of N increases, the computational time increases for both solvers. However, our proposed fast solver is less computationally expensive than the conventional direct solver in all cases. In particular, when N is small (e.g., N = 100), our fast solver achieves 166× runtime speed-up over the direct solver. Note that such a reduction in computational time is extremely important to our proposed

test application in real time.

4.2 Pre-test Analysis

The objective of pre-test analysis is to determine a set of predictable test items and the number of dies (i.e., N) that should be measured for these test items. We first physically measure all dies from the first wafer. Next, based on the measurement data, we perform linear search to find the optimal value of N with minimal test cost, as discussed in Section 3.1.B. In this example, the pre-defined target for both escape rate and yield loss is set to 0.5%. Since the proposed test cost reduction is applied to wafer probe measurement, a relatively large escape rate is allowed.

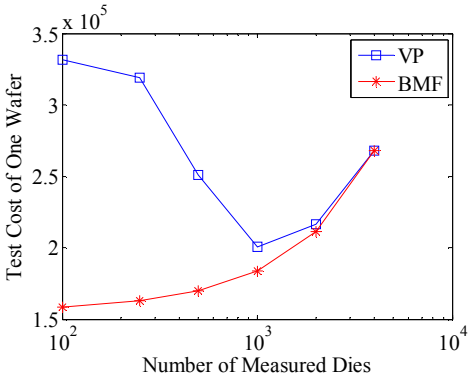

Figure 3. Test cost of a wafer, evaluated by the total number of measurements, is shown as a function of the number of dies (i.e., N) that should be measured for each predictable test item.

Figure 3 shows the test cost of a wafer, evaluated by the total number of measurements, as a function of N. Studying Figure 3 reveals two important observations. First, the test cost is not necessarily a monotonic function of N, as shown by the blue curve for VP. When N is too small, VP considers many test items as "unpredictable" and all dies must be measured for these test items. Hence, the test cost is large. On the other hand, if N is too large, a large number of dies must be measured even for predictable test items and, therefore, the test cost is large as well. These observations explain the reason why linear search is required to find the optimal value of N.

Second, for a given N, since BMF can capture the spatial variation more accurately than VP, more test items are classified as "predictable" by BMF than VP. As a result, the test cost of BMF is less than that of VP for any given N, as shown in Figure 3. Table 2 further shows the optimal values of N and the resulting numbers of predictable test items for VP and BMF, respectively.

Table 2. Pre-test analysis results for VP- and BMF-based test flow

	VP	BMF
Pre-defined Target for Escape Rate	5×10^{-3}	
Pre-defined Target for Yield Loss	5×10^{-3}	
Number of Measured Dies (N)	1000	100
Number of Predictable Test Items	33	34

4.3 Test Application

Table 3. Comparison of VP and BMF for test cost reduction

	Full	VP	BMF
Overall Test Cost	6.0×10^{7}	4.5×10^{7}	4.0×10^{7}
Test Cost Reduction	—	1.3×	1.5×
Escape Rate	—	2.5×10^{-5}	3.1×10^{-5}
Yield Loss	—	4.1×10^{-3}	1.6×10^{-3}

After pre-test analysis is complete, we apply the proposed test flow to all other wafers. Figure 4 shows the total number of dies that should be measured for each test item across all wafers. Here, three different cases are studied: (i) without test cost reduction (Full), (ii) with test cost reduction by VP (VP), and (iii) with test cost reduction by BMF (BMF). Note that both VP and BMF achieve significant reduction in test cost. In addition, the test cost of BMF is even lower than that of VP, since BMF needs to measure less dies for most test items. Table 3 summarizes the overall test cost (i.e., the total number of measurements across all wafers), the escape rate and the yield loss for VP and BMF. Based on Table 3, BMF achieves 1.125× reduction in test cost and 2.6× reduction in yield loss over VP in this example.

5. Conclusions

In this paper, we propose a novel statistical method, referred to as Bayesian model fusion (BMF), to accurately model the spatial variation at wafer level. The key idea is to exploit the fact that a large number of wafers of the same circuit design often share a similar spatial pattern. Hence, it is possible to borrow the measurement data from one wafer to model the spatial variation of other wafers. Several important tools (e.g., fast numerical solver) are developed to facilitate an efficient implementation of BMF for practical applications.

We further propose a new test methodology that

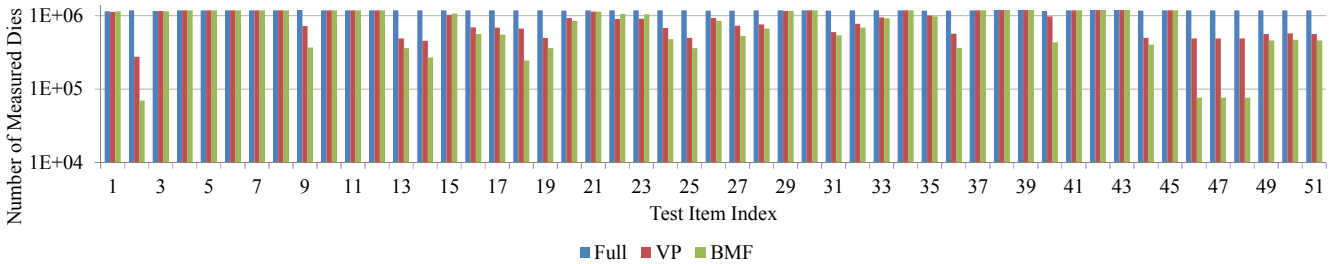

Figure 4. The total number of measured dies across all wafers is shown for each test item without test cost reduction (Full), with test cost reduction by VP (VP), and with test cost reduction by BMF (BMF).

incorporates BMF for test cost reduction. The proposed test flow consists of two phases: (i) pre-test analysis, and (ii) test application. A statistical model based on normal distribution is derived to closely monitor the escape rate and yield loss in order to guarantee the quality of the proposed test method. As is demonstrated by the wafer probe measurement data of an industrial RF transceiver, BMF achieves 1.125× reduction in test cost and 2.6× reduction in yield loss over the conventional approach based on VP. The reduction on test cost and/or yield loss could be more pronounced, if the spatial patterns of multiple wafers are closer to each other.

Finally, it is important to mention that several existing techniques (e.g., MVP [17]) have been proposed in the literature for spatial variation modeling over multiple wafers. However, unlike these conventional approaches that must first collect the measurement data from multiple wafers before modeling their spatial patterns, BMF allows us to extract the spatial pattern of a wafer as soon as the measurement data of the single wafer is available. By using BMF, we can test one wafer at a time, instead of waiting for the measurement data collected from multiple wafers. For this reason, BMF is preferred over the conventional methods for our proposed application of test cost reduction.

6. Acknowledgements

Support for this research was provided by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the Carnegie Mellon Portugal Program. This work was also supported by the National Science Foundation under contract CCF–1316363.

7. References

[1] K. Cheng and H. Chang, "Recent advances in analog, mixed-signal, and RF testing," *IPSJ Trans. on System LSI Design Methodology*, vol. 3, pp. 19-46, Feb. 2010.

[2] K. Arabi, "Mixed-signal test impact to SoC commercialization," *IEEE VTS*, 2010.

[3] P. Variyam, S. Cherubal and A. Chatterjee, "Prediction of analog performance parameters using fast transient testing," *IEEE Trans. on CAD*, vol. 21, no. 3, pp. 349-361, Feb. 2002.

[4] S. Biswas and R. Blanton, "Statistical test compaction using binary decision trees," *IEEE Design & Test of Computers*, vol. 23, no. 6, pp. 452-462, Jun. 2006.

[5] H. Stratigopoulos, P. Drineas, M. Slamani and Y. Makris, "Non-RF to RF test correlation using learning machines: A case study," *IEEE VTS*, 2007.

[6] R. Voorakaranam, S. Akbay, S. Bhattacharya, S. Cherubal and A. Chatterjee, "Signature testing of analog and RF circuits: algorithms and methodology," *IEEE Trans. on CAS – I*, vol. 54, no. 5, pp. 1018-1031, May. 2007.

[7] M. Chen and A. Orailoglu, "Test cost minimization through adaptive test development," *IEEE ICCD*, 2008.

[8] H. Stratigopoulos and Y. Makris, "Error moderation in low-cost machine learning-based analog/RF testing," *IEEE Trans. on CAD*, vol. 27, no. 2, pp. 339-351, Feb. 2008.

[9] D. Mannath, D. Webster, V. Montano-Martinez, D. Cohen, S. Kush, T. Ganesan and A. Sontakke, "Structural approach for built-in tests in RF devices," *IEEE ITC*, 2010.

[10] H. Stratigopoulos, P. Drineas, M. Slamani and Y. Makris, "RF specification test compaction using learning machines," *IEEE Trans. on VLSI*, vol. 18, no. 6, pp. 998-1002, Jun. 2010.

[11] H. Ayari, F. Azais, S. Bernard, M. Comte, M. Renovell, V. Kerzerho, O. Potin and C. Kelma, "Smart selection of indirect parameters for DC-based alternate RF IC testing," *IEEE VTS*, 2012.

[12] N. Akkouche, S. Mir, E. Simeu and M. Slamani, "Analog/RF test ordering in the early stages of production testing," *IEEE VTS*, 2012.

[13] H. Ayari, F. Azais, S. Bernard, M. Comte, V. Kerzerho, O. Potin and M. Renovell, "Making predictive analog/RF alternate test strategy independent of training set size," *IEEE ITC*, 2012.

[14] E. Yilmaz, S. Ozev and K. Butler, "Per-device adaptive test for analog/RF circuits using entropy-based process monitoring," *IEEE Trans. on VLSI*, vol. 21, no. 6, pp. 1116-1128, Jun. 2013.

[15] X. Li, R. Rutenbar and R. Blanton, "Virtual probe: a statistically optimal framework for minimum-cost silicon characterization of nanoscale integrated circuits," *IEEE ICCAD*, pp. 433-440, 2009.

[16] W. Zhang, X. Li and R. Rutenbar, "Bayesian virtual probe: minimizing variation characterization cost for nanoscale IC technologies via Bayesian inference," *IEEE DAC*, pp. 262-267, 2010.

[17] W. Zhang, X. Li, E. Acar, F. Liu and R. Rutenbar, "Multi-wafer virtual probe: minimum-cost variation characterization by exploring wafer-to-wafer correlation," *IEEE ICCAD*, pp. 47-54, 2010.

[18] H. Chang, K. Cheng, W. Zhang, X. Li and K. Butler, "Test cost reduction through performance prediction using virtual probe," *IEEE ITC*, 2011.

[19] W. Zhang, X. Li, F. Liu, E. Acar, R. Rutenbar and R. Blanton, "Virtual probe: a statistical framework for low-cost silicon characterization of nanoscale integrated circuits," *IEEE Trans. on CAD*, vol. 30, no. 12, pp. 1814-1827, Dec. 2011.

[20] C. Hsu, F. Lin, K. Cheng, W. Zhang, X. Li, J. Carulli and K. Butler, "Test data analytics - exploring spatial and test-item correlations in production test data," *IEEE ITC*, 2013.

[21] N. Kupp, K. Huang, J. Carulli and Y. Makris, "Spatial estimation of wafer measurement parameters using Gaussian process models," *IEEE ITC*, 2012.

[22] N. Kupp, K. Huang, J. Carulli and Y. Makris, "Spatial correlation modeling for probe test cost reduction in RF devices," *IEEE ICCAD*, pp. 23-29, 2012.

[23] K. Huang, N. Kupp, J. Carulli and Y. Makris, "Handling discontinuous effects in modeling spatial correlation of wafer-level analog/RF tests," *IEEE DATE*, pp. 553-558, 2013.

[24] K. Huang, N. Kupp, J. Carulli and Y. Makris, "On combining alternate test with spatial correlation modeling in analog/RF ICs," *IEEE ETS*, 2013.

[25] E. Yilmaz and S. Ozev, "Adaptive multi-site testing for analog/mixed-signal circuits incorporating neighborhood information," *IEEE ETS*, 2012.

[26] R. Gonzalez and R. Woods, *Digital Image Processing*, Prentice Hall, 2007.

[27] X. Li, W. Zhang, F. Wang, S. Sun and C. Gu, "Efficient parametric yield estimation of analog/mixed-signal circuits via Bayesian model fusion," *IEEE ICCAD*, pp. 627-634, 2012.

[28] F. Wang, W. Zhang, S. Sun, X. Li and C. Gu, "Bayesian model fusion: large-scale performance modeling of analog and mixed-signal circuits by reusing early-stage data," *IEEE DAC*, 2013.

[29] C. Gu, E. Chiprout and X. Li, "Efficient moment estimation with extremely small sample size via Bayesian inference for analog/mixed-signal validation," *IEEE DAC*, 2013.

[30] S. Sun, F. Wang, S. Yaldiz, X. Li, L. Pileggi, A. Natarajan, M. Ferriss, J. Plouchart, B. Sadhu, B. Parker, A. Valdes-Garcia, M. Sanduleanu, J. Tierno and D. Friedman, "Indirect performance sensing for on-chip analog self-healing via Bayesian model fusion," *IEEE CICC*, 2013.

[31] X. Li, F. Wang, S. Sun and C. Gu, "Bayesian model fusion: a statistical framework for efficient pre-silicon validation and post-silicon tuning of complex analog and mixed-signal circuits," *IEEE ICCAD*, pp. 795-802, 2013.

[32] G. Golub and C. Loan, *Matrix Computations*, Johns Hopkins Univ. Press, 1996.

[33] C. Bishop, *Pattern Recognition and Machine Learning*, Prentice Hall, 2007.

[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.

[35] Z. Drezner and G. Wesolowsky, "On the computation of the bivariate normal integral," *Journal of Statistical Computation and Simulation*, vol. 35, pp. 101-107, 1990.