Test Cost Reduction Through Performance Prediction Using Virtual Probe

Hsiu-Ming (Sherman) Chang and Kwang-Ting (Tim) Cheng
University of California, Santa Barbara
{sherman, timcheng}@ece.ucsb.edu

Wangyang Zhang and Xin Li
Carnegie Mellon University
{wangyan1, xinli}@ece.cmu.edu

Kenneth M. Butler
Texas Instruments
kenb@ti.com

Abstract — The virtual probe (VP) technique, based on recent breakthroughs in compressed sensing, has demonstrated its ability for accurate prediction of spatial variations from a small set of measurement data. In this paper, we explore its application to cost reduction of production testing. For a number of test items, the measurement data from a small subset of chips can be used to accurately predict the performance of other chips on the same wafer without explicit measurement. Depending on their statistical characteristics, test items can be classified into three categories: highly predictable, predictable, and un-predictable. A case study of an industrial RF radio transceiver with more than 50 production test items shows that a good fraction of these test items (39 out of 51 items) are predictable or highly predictable. In this example, the 3σ error of VP prediction is less than 12% for predictable or highly predictable test items. Applying the VP technique can on average replace 59% of test measurement by prediction and, consequently, reduce the overall test time by 57.6%.

1. Introduction

Testing cost is a significant component of the overall product cost for modern integrated circuits. In particular, testing mixed-signal and RF components in a system on chip (SOC) to examine their conformance to specifications [1] could account for up to 70% of the overall test cost of a mixed-signal SoC [2]. In addition to random defects and systematic failures that could result in defective devices, parametric variations in circuit/device/process parameters (such as gate length, dopant concentration, and metal thickness) cause deviations in device performances (such as gain, power, and bandwidth) and, hence, also lead to yield loss. As technology continues to scale, yield loss due to systematic failures and parametric variations become increasingly significant. While such failures are difficult to screen, in contrast to random defects, they often present strong spatial die-to-die correlations at the wafer level. Hence, statistical methods exploring such spatial correlations for performance prediction offer a promising direction to achieve test cost reduction, since they allow us to replace a portion of physical measurements by predicted performance values that are derived from the measurement data of a small subset of chips on a wafer.

Several learning-based methods have been proposed to characterize the parametric variations for performance prediction and to explore the tradeoff between prediction accuracy and test cost [3-6]. For example, the *alternate test* framework [5-6] attempts to predict circuit performances based on a set of signatures captured from cheaper and simpler test setups and measurements on the device-under-test (DUT). The key assumption behind alternate testing is that the DUT's signature values and performance values are strongly correlated, as they both are affected by the same parametric variations. Therefore, the DUT's performance values can be predicted from its signature values once the correlations between them are accurately learned. Estimating such correlations usually requires a model training process. Since the correlation models could differ from lot to lot, they must be trained separately for different manufacturing batches.

The *virtual probe* (VP) technique, proposed recently in [7-9], can be employed to reduce silicon characterization cost. The basic idea is to randomly [9] or iteratively [8] select only a small subset of test structures in the silicon wafer for physical measurement, and the parametric variations at other locations on the same wafer can be predicted using a statistical algorithm. In other words, VP aims to construct the spatial variations of the entire wafer based on as few measurements as possible. In addition, the multi-wafer virtual probe (MVP) algorithm developed in [7] further improves the prediction accuracy by exploring the strong correlations among different wafers within the same lot.

VP demands significantly less effort for test development than the learning-based methods, such as alternate testing. First, VP does not require any design modification, nor any additional design-for-testability (DfT) circuitry. Moreover, there is no requirement for developing special test stimuli for performance measurements and prediction. In addition, VP does not require model training and thus is more scalable and applicable to cases where measurement data are limited or non-stationary due to low-volume manufacturing or using multiple equipment sets.

In [7-9], the effectiveness of VP was demonstrated through the prediction of circuit delay and leakage – including flush delay, ring oscillator frequency, leakage power, and leakage current – of test structures and digital circuits. In this paper, we further investigate the possibility of applying VP to production testing of mixed-signal/RF circuits. Towards this goal, several key issues must be carefully addressed. First, due to the high design complexity

978-1-4577-0152-8/11/$26.00

©2011 IEEE

and the sophisticated interaction among different building blocks in both time and frequency domains, the performances of the mixed-signal/RF DUTs could exhibit substantially more complex characteristics in performance variability than those of test structures and digital circuits. Production testing of a typical mixed-signal/RF device often consists of dozens of test items, ranging from DC/AC parameters to time/frequency-domain specifications. Some of these test items carry stronger spatial correlations than the others. Those items with strong (weak) spatial correlations are expected to have high (low) prediction accuracy. Hence, classifying all test items based on their degree of predictability is of great importance.

Because VP utilizes spatial correlations among dies for performance prediction, random defects that do not exhibit spatial correlations cannot be predicted. Screening of defective chips resulting from such random failures by prediction without any explicit testing will be impossible. However, random-defect-induced defective chips often violate multiple specs and thus can be screened by explicit testing using a carefully selected and low-cost subset of test items. Only those chips passing this subset of test items are considered for VP-based prediction. One last issue to point out is that the prediction accuracy used for product screening should be evaluated based on 3σ or even 6σ errors, not simply the average prediction errors considered in [7-9].

The main contribution of this paper is a unique proposal which uses VP to minimize the time and cost for testing mixed-signal/RF circuits and a case study based on the real test data of an industrial dual-radio RF transceiver. The proposed method requires screening of defective chips with catastrophic failures prior to applying VP-based prediction. Our study provides many useful insights to answering the question of how to sample the spatial locations for explicit measurement so as to minimize the 3σ/6σ prediction errors for those test items without explicit measurement. Because each test item has different degrees of spatial correlations, we classify all test items, based on the accuracy of VP prediction, into three categories: un-predictable, predictable, and highly predictable. Based on the silicon measurement data, we demonstrate that a good fraction of an industrial dual-radio RF transceiver's test items are highly predictable. It, in turn, validates the efficacy of the proposed method for test cost reduction.

The rest of the paper is organized as follows. Section 2 briefly summarizes the technical background for the VP technique. Section 3 describes the details of our proposed test methodology, followed by a case study of an industrial RF radio chip in Section 4. Section 5 concludes the paper.

2. Background: Virtual Probe

The key idea of virtual probe (VP) [7-9] is to test a subset of chips at selected locations on a wafer and, then, using a statistical algorithm, predict the circuit performance of other chips, instead of explicitly testing them. Such performance prediction is made possible by carefully modeling the wafer-level performance variation in the (spatial) frequency domain. In this section, we briefly review the VP techniques that were proposed in [7-9].

2.1. Mathematical Formulation

Consider a performance metric g (e.g., speed, gain, etc.) that should be measured for all chips from L different wafers. For each wafer, the measured performance g is most likely different from chip to chip due to process variations. Such a wafer-level spatial variation can be represented as a two-dimensional function $g_{(l)}(x, y)$ ($l = 1,2,\ldots L$), where l denotes the l-th wafer and x and y represent the coordinate of a chip within the wafer. Without loss of generality, x and y can be labeled as integer numbers: $x \in \{1,2,...,P\}$ and $y \in \{1,2,...,Q\}$.

Mathematically, the spatial variation $g_{(l)}(x, y)$ can be mapped to the frequency domain by a two-dimensional linear transform such as discrete cosine transform (DCT) [7-9]:

$$G_{(l)}(u,v) = \sum_{x=1}^{P}\sum_{y=1}^{Q} \alpha_u \cdot \beta_v \cdot g_{(l)}(x,y) \cdot \cos\frac{\pi(2x-1)\cdot(u-1)}{2\cdot P} \cdot \cos\frac{\pi(2y-1)\cdot(v-1)}{2\cdot Q} \quad (l = 1,2,\cdots,L) \tag{1}$$

where

$$\alpha_u = \begin{cases} \sqrt{1/P} & (u=1) \\ \sqrt{2/P} & (2 \le u \le P) \end{cases} \tag{2}$$

$$\beta_v = \begin{cases} \sqrt{1/Q} & (v=1) \\ \sqrt{2/Q} & (2 \le v \le Q) \end{cases}. \tag{3}$$

In (1), $\{G_{(l)}(u, v);\ u = 1,2,...,P,\ v = 1,2,...,Q\}$ represents the DCT coefficients (i.e., the frequency-domain components of the spatial variation). Equivalently, the performance values $\{g_{(l)}(x, y);\ x = 1,2,...,P,\ y = 1,2,...,Q\}$ can be represented as the linear combinations of $\{G_{(l)}(u, v);\ u = 1,2,...,P,\ v = 1,2,...,Q\}$ by inverse discrete cosine transform (IDCT):

$$g_{(l)}(x,y) = \sum_{u=1}^{P}\sum_{v=1}^{Q} \alpha_u \cdot \beta_v \cdot G_{(l)}(u,v) \cdot \cos\frac{\pi(2x-1)(u-1)}{2\cdot P} \cdot \cos\frac{\pi(2y-1)(v-1)}{2\cdot Q} \quad (l = 1,2,\cdots,L). \tag{4}$$

VP [7-9] aims to measure a small number of chips from a wafer and then predict the performance value $g_{(l)}(x, y)$ of other chips on the same wafer. As such, the time and cost associated with the performance measurement can be reduced. Towards this goal, the following linear equations are formulated:

$$A_{(l)} \cdot \eta_{(l)} = B_{(l)} \tag{5}$$

where

$$A_{(l)} = \begin{bmatrix} A_{(l),1,1,1} & A_{(l),1,1,2} & \cdots & A_{(l),1,P,Q} \\ A_{(l),2,1,1} & A_{(l),2,1,2} & \cdots & A_{(l),2,P,Q} \\ \vdots & \vdots & \vdots & \vdots \\ A_{(l),M_l,1,1} & A_{(l),M_l,1,2} & \cdots & A_{(l),M_l,P,Q} \end{bmatrix} \tag{6}$$

$$A_{(l),m,u,v} = \alpha_u \cdot \beta_v \cdot \cos\frac{\pi(2x_{(l),m}-1)\cdot(u-1)}{2\cdot P} \cdot \cos\frac{\pi(2y_{(l),m}-1)\cdot(v-1)}{2\cdot Q} \quad (7)$$

$$\eta_{(l)} = [G_{(l)}(1,1) \quad \cdots \quad G_{(l)}(P,Q)]^T \quad (8)$$

$$B_{(l)} = [g_{(l)}(x_{(l),1}, y_{(l),1}) \quad \cdots \quad g_{(l)}(x_{(l),M_l}, y_{(l),M_l})]^T. \quad (9)$$

In (9), the vector $B_{(l)}$ contains the performance values measured from $M_{(l)}$ different chips $\{(x_{(l),m}, y_{(l),m}); m = 1,2,\ldots,M_{(l)}\}$ of the l-th wafer. Once the linear equations in (5) are solved to determine the solution $\eta_{(l)}$, i.e., the DCT coefficients $\{G_{(l)}(u, v); u = 1,2,...,P, v = 1,2,...,Q\}$, the performance value $g_{(l)}(x, y)$ can be calculated for any spatial location (x, y) by the IDCT in (4).

Solving the linear equations in (5), however, is not trivial. Note that the number of equations, i.e., $M_{(l)}$, is significantly less than the number of unknowns, i.e., the product of P and Q. In other words, the linear equations in (5) are profoundly under-determined. In this case, the solution $\eta_{(l)}$ is not unique, unless additional constraints are added.

To uniquely determine $\eta_{(l)}$, VP further assumes that $\eta_{(l)}$ is sparse. Namely, a large number of DCT coefficients are close to zero, but we do not know the exact locations of these zeros. In general, if the performance variation $\{g_{(l)}(x, y); x = 1,2,...,P, y = 1,2,...,Q\}$ presents a spatial pattern, i.e., the variation is spatially correlated, the vector $\eta_{(l)}$ that contains the corresponding DCT coefficients is sparse. This sparsity has been observed in many image processing tasks. As was demonstrated by several industrial examples in [7-9], such a sparseness assumption is also valid for some data collected from real silicon with advanced VLSI technologies.

Given the aforementioned sparseness assumption, the solution $\eta_{(l)}$ of (5) can be uniquely determined by solving the following optimization [9]:

$$\begin{aligned} \underset{\eta_{(l)}}{\text{minimize}} \quad & \|\eta_{(l)}\|_0 \\ \text{subject to} \quad & A_{(l)} \cdot \eta_{(l)} = B_{(l)} \end{aligned} \quad (10)$$

where $\|\eta_{(l)}\|_0$ stands for the L_0-norm of $\eta_{(l)}$, i.e., the number of non-zeros in $\eta_{(l)}$. The optimization in (10) attempts to minimize the number of non-zeros in $\eta_{(l)}$, while satisfying the linear equation $A_{(l)} \cdot \eta_{(l)} = B_{(l)}$. Hence, it results in a unique solution $\eta_{(l)}$ that is as sparse as possible.

The optimization problem in (10) is NP hard and, hence, is extremely difficult to solve [9]. A more efficient technique to find sparse solutions is based on L_1-norm regularization – a relaxed version of L_0-norm [9]:

$$\begin{aligned} \underset{\eta_{(l)}}{\text{minimize}} \quad & \|\eta_{(l)}\|_1 \\ \text{subject to} \quad & A_{(l)} \cdot \eta_{(l)} = B_{(l)} \end{aligned} \quad (11)$$

where $\|\eta_{(l)}\|_1$ denotes the L_1-norm of $\eta_{(l)}$, i.e., the summation of the absolute value of all elements in $\eta_{(l)}$. The L_1-norm problem in (11) can be re-formulated as a linear programming problem and solved efficiently [9].

2.2. Bayesian Virtual Probe

Bayesian virtual probe (BVP) [8] can be conceptually viewed as an extended version of the aforementioned VP technique. It identifies the optimal spatial locations where silicon chips should be tested to predict the spatial variation of the entire wafer with maximum accuracy. Finding the optimal sampling locations, however, is not trivial, since they strongly depend on the spatial pattern of process variations. Namely, the optimal sampling locations are different from process to process, from wafer to wafer, and from chip to chip. It is impossible to come up with a fixed set of sampling locations that are optimal for all cases. Instead, the best sampling locations must be adaptively "learned" in real time.

For this reason, BVP starts from a small set of measurement data by testing very few chips on a wafer. It applies VP to predict the wafer-level spatial variation and then estimates the resulting prediction error based on Bayes' theorem. Next, the optimal sampling locations are found by information theory to collect additional test data and improve prediction accuracy. The aforementioned two steps (i.e., error estimation and optimal sampling) are repeatedly applied until the prediction accuracy is sufficiently high.

3. VP-based Test Methodology

3.1. Predictability of Performance Parameters

One important assumption that VP is rooted upon – the DCT coefficients of the spatial distribution are sparse – is true for a majority, but not all, of the performance metrics. For example, Vth, the threshold voltage of a transistor is extremely sensitive to the dopant variations and could have large and unpredictable variations across a die, not to mention the Vth variations for neighboring dies. Such performance metrics/parameters will therefore be unpredictable using VP. On the other hand, other performance parameters, such as the longest propagation delay of the chip, may have strong die-to-die spatial correlations across a wafer, and thus would be ideal for applying the VP technique. Therefore, depending on the strength of its wafer-level spatial correlations, a performance parameter may or may not be predictable using VP.

It is, however, impossible to know the spatial distribution of the performance parameters *a priori*. In practice, we can fully characterize one wafer and study its spatial variations by conducting a pre-test analysis prior to applying the actual testing procedure. In this analysis, we can determine the predictability of a performance metric using a cross-validation technique: we can repeatedly run VP using a small randomly-selected subset of the measurement data and compare the prediction of another sample set with their actual measurement data. This cross-validation result could reveal the degree of predictability.

Based on this validation process, we can classify each test item (which tests a specific performance metric) into three categories: *un-predictable*, *predictable*, and *highly-predictable*. The classification could be based on the required sample size for VP to achieve a satisfactory level of prediction accuracy. For example, an unpredictable parameter might require making explicit measurement for more than 90% of the chips in order to predict, with sufficient accuracy, the performance of the other 10% of the chips. On the other hand, a highly-predictable parameter may only require explicit measurement of 10% of the chips in order to predict the other 90% of the chips with high accuracy. For those highly-predictable parameters identified and validated in this way, we can then confidently employ the VP technique to save test time and cost.

VP and BVP achieve different degrees of efficiency for highly predictable and predictable parameters. For *highly predictable* parameters, VP-based prediction is sufficiently accurate and, hence, further applying BVP only achieves limited benefit. On the other hand, for parameters in the *predictable* category, BVP can reduce the number of required samples while achieving the same prediction accuracy as VP.

3.2. Pre-test Analysis Using VP

The goal of pre-test analysis is to identify a subset of locations for each test item to be predicted, instead of explicitly measured in testing, to achieve minimum test cost without compromising test quality. This goal can be achieved by conducting a thorough analysis and cross-validation on one or several wafers and then applying the findings to develop an optimal test strategy for other wafers. The aforementioned analysis and cross-validation include: 1) classifying each test item into one of the three categories, and 2) for each highly-predictable or predictable test item, determining the number of required die samples, and their locations on the wafer, for explicit test measurement.

We first use an iterative sampling-and-validation method on one wafer to determine the number of samples required to achieve a satisfactory level of prediction accuracy using VP. For this wafer, we test every die for every test item and record their measurement data for analysis. In this iterative analysis, we first sample a fraction of the dies, say 10%, at randomly selected locations for VP analysis and use another independent subset, say 10%, of random samples, to evaluate the prediction error. If the 3σ or 6σ of the prediction error for the test-item-under-analysis is less than a target threshold (could be different for each test item) the test item is classified as highly predictable, and the sample count as well as the corresponding locations of the selected random samples will be used for production testing of other wafers for this test item. To reduce the variability of random sampling for the proposed iterative sampling-and-validation method, we can repeat the aforementioned procedure for multiple times, each of which chooses different locations for the training and testing sets. For each test item, the sampling locations that result in the smallest prediction error will be the preferred choices used for production testing. Such an approach allows us to identify a set of sampling locations that yield less prediction error than simple random sampling and, hence, reduce the testing cost. Note that determining the globally optimal sampling locations is an extremely difficult, if not impossible, task because it requires to exhaustively search all possible sampling locations and their combinations. However, we can further apply BVP as an efficient heuristic algorithm to find "good" sampling locations. More details on BVP are described in Section 3.3.

If the prediction error exceeds the target threshold, the sample size is then increased to, say 20%, for VP analysis followed by cross-validation. This iterative process continues until the prediction error falls below the target threshold. Such an iterative process is also repeated for each test item. The target threshold for a test item is determined based on the trade-offs between the cost of explicitly testing (i.e., physically measuring the test item) and the cost of test escape (i.e., due to the prediction error of VP). In addition, we could use different thresholds for a test item for different applications. For example, we might want to set a smaller threshold for the linearity spec of analog-to-digital converters (ADCs) used in imaging than in audio systems because the former requires a higher linearity while the later has a more demanding spec on harmonics.

Test items that can be predicted from a small sample size are classified as highly-predictable. For these items, the prediction error stays relatively constant even if the sample size is increased by 2-4x, implying that the spatial correlations have already been captured accurately with a small sample size. For these test items, VP with a small sample size is sufficient to achieve high prediction accuracy. In our experiment, we classify a test item as highly-predictable, if the required sample size is less than 20% of available chips.

Test items which cannot reach the target prediction error (unless using more than 90% samples) are unpredictable and will need to be measured for every die on every wafer. Test items that are neither un-predictable nor highly predictable are classified as predictable.

For predictable and highly predictable test items, we can apply BVP to determine the best sampling locations. More details on the topic can be found in Sections 3.3.

Finally, it is worth mentioning that, from time to time, we can sample additional dies for measurement to revalidate the prediction accuracy. In the event of observing an increased prediction error (e.g., a non-trivial number of dies with a prediction error greater than the 3σ bound obtained in the pre-test analysis), we will re-run the pre-test analysis to update the test item selection. This re-evaluation process could be applied to one wafer for each lot on a regular basis or when there are changes in the fabrication environment.

3.3. Using BVP for Reducing Prediction Errors

Table 1: σ, test time, and categories of representative test items.

Test Item Number	Description	Test Time Estimation*	Estimated 3σ Prediction Error Requirement	Category
1	Bit Error Rate	10	6%	Un-predictable
7	Receiver voltage	3	12%	Highly Predictable
11	Receiver current 1	3	12%	Highly Predictable
14	Receiver current 2	3	12%	Predictable
33	Power measurement 1	7	10%	Highly Predictable
37	Power measurement 2	7	10%	Predictable
50	Standby current	3	12%	Un-predictable

* Note: Test time estimation is rated scale 1 to 10, with the 10 being the most time consuming test item.

BVP adaptively selects optimal locations for explicit test that helps achieve better prediction accuracy than VP without increasing the sample size. For a test item, suppose that a random sample set *S* is chosen by VP to meet a given target threshold. BVP chooses a subset of *S*, for example half of *S*, as the initial number of samples and iteratively selects additional samples until the target accuracy is reached. We can repeat this iterative process multiple times, each with a different initial sample size. The sample size and the corresponding locations of the best result among these multiple runs are then used in the production test.

Note that BVP is only applied during the pre-test analysis to determine the best sampling locations for each test item. In production testing, test measurement will be performed at the pre-determined testing locations for each test item, and adaptive sampling is not being applied. It is also important to note that different dies may be associated with different test items for explicit testing.

3.4. Test Application

The pre-test analysis yields a list of test items for each die. Some dies will be tested with more test items than others. Other test items that are not applied/measured for a die will be predicted accurately by VP. This die-location-dependent test plan is incorporated into the test program for actual test application.

Since the performance values of defective chips with catastrophic failures will not follow any spatial correlation, they should be screened out prior to applying the VP-based test methodology. Because these defective chips usually fail multiple performance parameters, we can apply simple, low-cost tests first to identify them. For example, gain and offset errors are good indicators of whether an ADC can perform basic conversions. These two test items should be performed first to filter out ADCs that have catastrophic failures caused by random defects. On the other hand, for phase-locked loops (PLLs), a simple frequency measurement can be used to determine whether it can successfully reach a clock generating state [10]. Any random defects that affect one or more building blocks within the loop will likely cause fail-to-lock condition. Other methods presented in the literature, such as the structural-based test approach for RF devices in [11] and the defect-oriented test methods for mixed-signal circuits in [12], can effectively detect catastrophic failures at a relatively low cost. We can apply these simple tests first in our test application and terminate the test process once a failure is detected.

The VP-based test methodology can be applied during both wafer probing and production test after packaging. In both cases, the locations of each die on the wafer are recorded so that VP can reconstruct the spatial variations of the wafer.

3.5. Summary of the proposed test methodology

In the pre-test analysis:

1) For each test item on a single wafer, use sample-and-validation method to determine the sample size for VP.
2) Determine the category of each test item based on the required sample size to achieve a target level of accuracy.
3) For predictable and highly predicable items, determine if BVP should be applied.
4) If BVP should be applied, use BVP to select the optimal sampling locations.

In the test application:

1) Apply simple tests to detect defective chips that have catastrophic failures.
2) Based on the pre-test analysis results, apply a die-location-dependent test plan, collect measurement data from one/multiple wafer(s), and run VP to predict the performance of those that are not explicitly measured.
3) If VP is used, for each test item, cross-validate the prediction accuracy for one wafer in a lot by comparing the predicted values with additional measurements from a subset of, say 10%, dies at randomly selected locations. If the 3σ error is greater than a target limit, re-run the pre-test analysis. Otherwise, conduct Step (2) for the rest of the wafers in the same lot.

4. Case Study of a Dual Radio RF Transceiver

We applied the VP-based test flow to the wafer probe measurement data of an industrial RF transceiver chip which includes two radios. The dataset consists of 175 wafers from 9 lots, and each wafer has over 6000 dies. We examined 51 test items including bit error rate (BER), power, current, and voltage measurements.

Table 1 lists several representative test items from the total of 51. In the table, a short description of each listed item is given. For each item, the estimated test time is

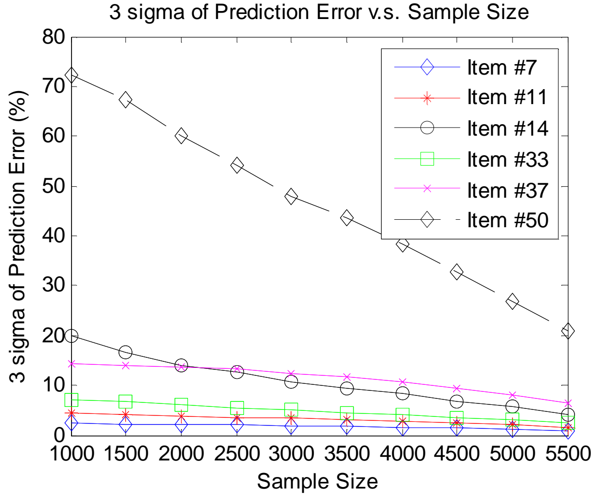

Figure 1: Sample-and-validation results show that 3σ of the prediction error decreases as the number of samples increases.

denoted on a scale from 1 to 10, where a value of 10 corresponds to the most time consuming test item. In addition, each test item has a unique requirement for the 3σ prediction error. In our experiment, we set a smaller error tolerance for those items that have longer test times, which reflects their more stringent test precision requirements. The category column reports the classification result of the corresponding test items after our pre-test analysis.

4.1. Pre-Test Analysis Result

Among the 51 test items, the BER and standby current are highly sensitive to independent random variations of a number of device and circuit parameters and thus are expected to have significant standard deviations and weak spatial correlations. Therefore, these types of test items are likely un-predictable. Other test items such as voltage, power, and current measurements have stronger spatial correlations and are likely to be highly predictable or predictable.

During the first step in the pre-test analysis process, we thoroughly analyze one of the wafers. For this device, among all the dies on the wafer being analyzed, a few die are screened, and the remaining dies (more than 6000 in total) are included in our pre-test analysis.

We start with 1,000 random samples, out of the remaining 6000+ dies, for our proposed sampling-and-validation process and increment an additional 500 samples for each subsequent iteration. In each iteration, for each die and for each predicted test item, we calculate the prediction error as:

$$PredictionError(x, y) = \left| \frac{g(x, y) - g^*(x, y)}{g(x, y)} \right|, \qquad (12)$$

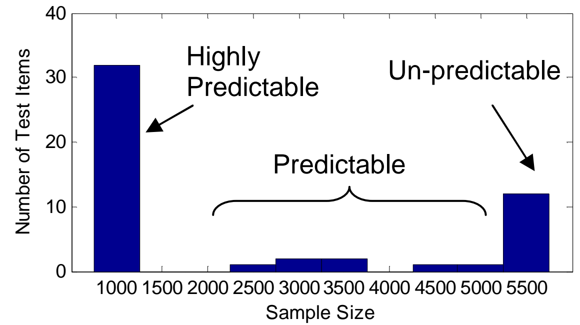

Figure 2: Histogram showing sample size for each test item.

where x, y are the locations of the die on the wafer, and $g(x, y)$ and $g^*(x, y)$ denote the actual and the predicted performance values respectively. For each test item, we compute the average, standard deviation (σ), and 3σ values of the *PredictionError* of all predicted dies.

Figure 1 shows the 3σ prediction error as a function of the number of samples for several test items listed in Table 1. Test item #1 (BER Test) is not plotted in Figure 1 because it has a prediction error around 700% that is significantly higher than the error of other items. Figure 1 shows that, as the sample size increases, VP could better estimate the spatial correlations, thereby reducing the prediction error for all test items. In addition, each test item requires different sample sizes to achieve the same prediction accuracy. For example, to make the 3σ prediction error lower than the target thresholds, test items #7, #11, #33 would require 1,000 samples, and test items #14 and #37 would need 3,000 and 4,500 samples, respectively. Accordingly to our classification, test items #7, #11, and #33 are highly predictable whereas test items #14 and #37 are predictable. On the other hand, test items #1 and #50 are unpredictable as they would need a large amount of samples to achieve the given target accuracy.

Figure 2 shows the histogram of sample sizes required for the 51 test items. For un-predictable items, we need to explicitly test all dies and, thus, the number of samples equals the total number of dies on a wafer. Among the 51 test items studied, our pre-test analysis classifies 32 as highly predictable, 7 as predictable, and 12 as un-predictable, which account for 62%, 14%, and 24% of the overall test items reviewed, respectively.

Figure 3 depicts the 3σ of the measurement values and the 3σ of the prediction errors for both the predictable and highly-predictable test items. Each circle represents a test item and the diagonal line represents the boundary where x and y axes have the same values. For this experiment, we randomly sampled 3,500 locations and predict the remaining 2500+ dies. It is clear that all the circles are at the right side of the diagonal line – which confirms that VP does capture the spatial correlation because it yields a 3σ prediction error much lower than the 3σ of the measurement data. In addition, for the same sampling size, test items with a

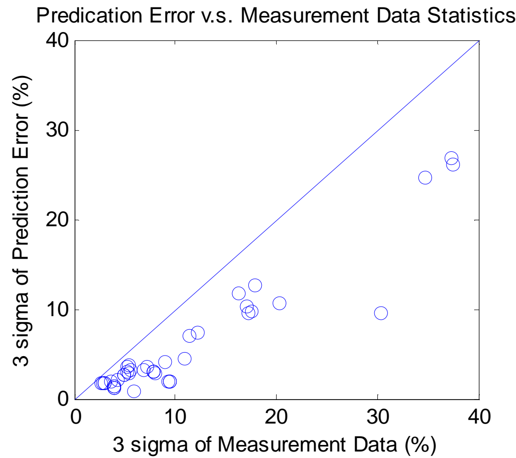

Figure 3: Comparison of the 3σ of the measurement data and the prediction error. The sample size are 3,500 for all test items.

smaller standard deviation of measurement values tend to have a smaller prediction error too.

The distribution of the measurement data for test item #33 (a power measurement), a highly predictable test item, is illustrated in Figure 4(a). The 3σ of the measurement data is 10.86%. We randomly selected 3,500 dies and used VP to replace the measurement of the remaining 2500+ dies. The histogram of the prediction error for all the dies on the wafer is drawn in Figure 4(b), where the average error is 2.11% and the 3σ error is 7.12%. The same experiment is conducted for test item #14 (a receiver current), a predictable test item. The results are shown in Figures 5(a) and (b). The 3σ values of the performance and the prediction error are 30.32% and 13.82%, respectively. Both results show that, regardless of the data distributions, VP can predict a great majority of the performance values with high accuracy.

Figures 4(c) and 5(c) illustrate the normalized measurement values of test items #33 and #14 on the wafer, respectively. While both test items have a similar spatial distribution, the variation of item #14 is larger. Therefore, given the same sample size, item #14 has a greater prediction error than item #33. On the other hand, Figure 6 plots the normalized measurement values of test item #1 (bit error rate, an un-predictable item) on the wafer. It shows that no spatial correlation exists. Therefore, it is impossible to predict its value using VP.

4.2. Evaluation of Test Application

In the pre-test analysis, for each die, we determine a list of test items for explicit testing. We then apply VP-based test method to all wafers. Test items without explicit measurement are predicted using VP. The maximum of the 3σ prediction error in each lot for predictable and highly predictable items is shown in Table 2.

The results in Table 2 show that, after pre-test analysis, the VP test method can be successfully applied to wafers in different lots to achieve a prediction error less than the target threshold for most test items. However, for test item #14, the maximum of the 3σ prediction errors is larger than the target threshold for lots 3, 6, 7 and 8. Such a large prediction error can be detected if we explicitly measure all

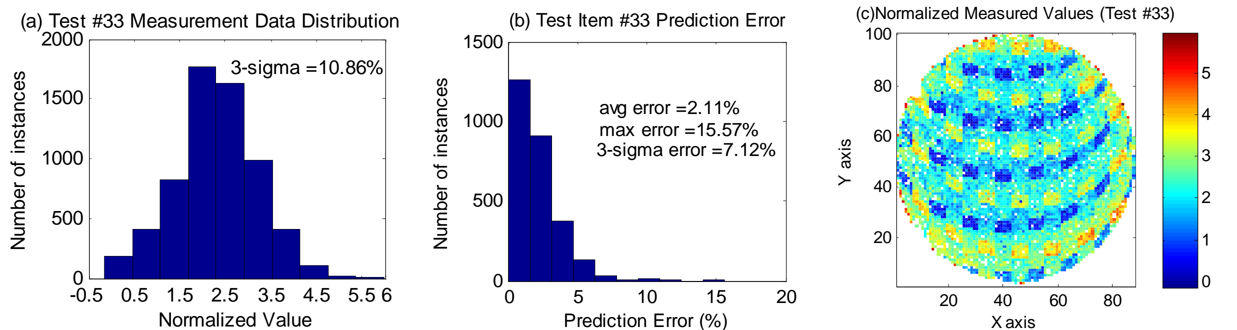

Figure 4: Histogram of the test item #33 and the prediction error calculated by Eq. (12) for all the chips on the same wafer

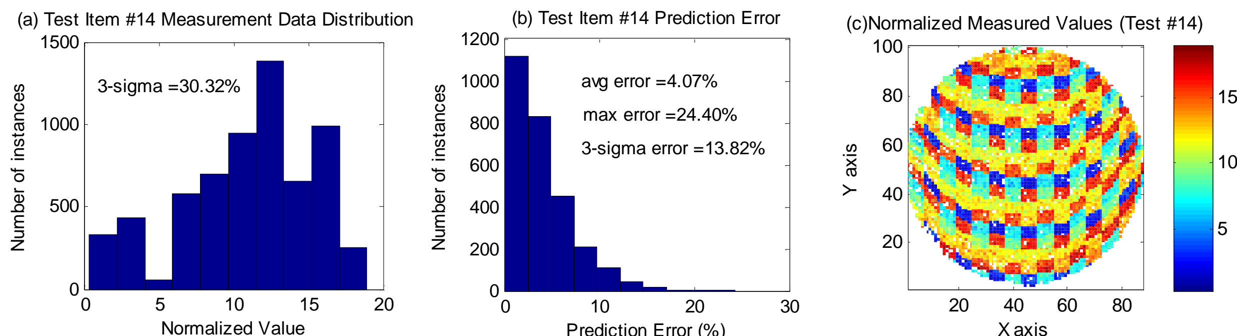

Figure 5: Histogram of the test item #14 and the prediction error calculated by Eq. (12) for all the chips on the same wafer.

Table2: The maximum 3σ prediction error for applying VP-based test method to wafers in different lots.

Test Item Number	Sample Size	Lot1	Lot2	Lot3	Lot4	Lot5	Lot6	Lot7	Lot8	Lot9
Number of Wafers		8	25	23	25	2	17	25	25	25
7	1,000	4.26%	4.70%	4.02%	4.77%	3.57%	3.74%	4.00%	4.25%	5.70%
11	1,000	6.03%	6.18%	6.04%	6.23%	5.34%	5.77%	5.80%	5.98%	6.62%
14	4,000	9.46%	10.72%	12.75%	10.00%	10.51%	16.71%	15.34%	16.62%	9.08%
33	1,000	8.38%	8.45%	8.51%	9.65%	7.50%	10.52%	8.75%	9.10%	9.27%
37	4,500	11.20%	11.48%	11.51%	11.43%	11.72%	11.05%	11.21%	11.69%	11.39%

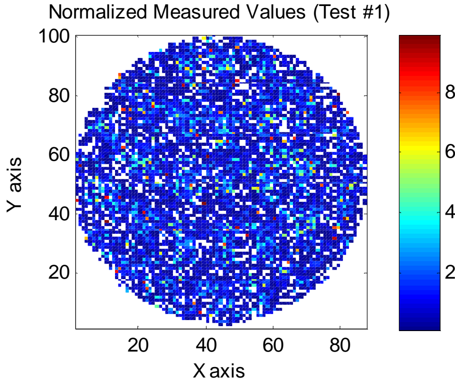

Figure 6: Measured value (normalized with a random number) for test item 1 (bit error rate) test.

test items for a subset of dies, say 10%, in each lot for cross-validation. If a large 3σ prediction error is observed (such as the case for lots 3, 6, 7 and 8), we should re-run the pre-test analysis. On the other hand, we can also re-run the pre-test analysis when a much smaller 3σ prediction error is detected. This means that further cost reduction may be possible for this lot. In either case, we re-run the pre-test analysis if inconsistency is observed.

Without VP, the total number of measurements for all test items per wafer would be the number of test items multiplied by the number of samples, which is quite large in our experiment. With VP, 59% of the measurements could be replaced by prediction. We further weight each item with their estimated test time to evaluate the savings in test time. For each die, the total normalized test time required for applying the 51 test items studied is 305 time-units. For the number of samples we used, the total test time would be approximately 1,900,000 time-units. With VP, the total normalized test time is reduced to approximately 82,000 time-units, yielding 2.36x speedup in test time.

We further make pass/fail decisions based on the VP-based test results. To accommodate the prediction errors, we add margins to the pass/fail threshold from 5% to 15%. Compared to the decisions made using conventional test results, we do not observe any test escape or yield loss from the 175 wafers in our experiment. However, extra measurement data are required to accurately estimate the actual escape rate because such results could be parameter and data dependent.

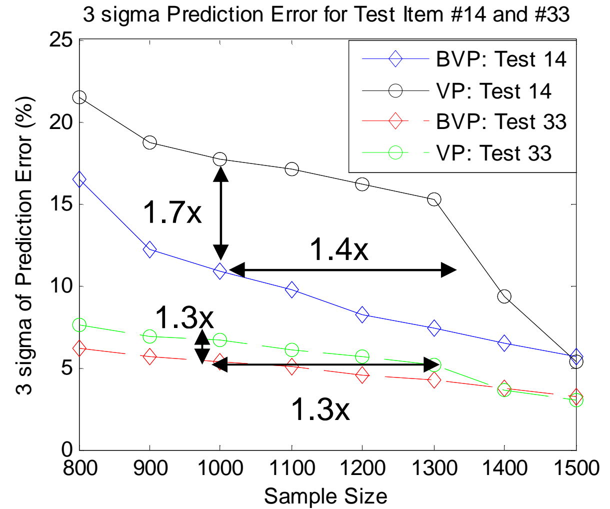

Figure 7: Comparison of the 3 sigma prediction errors for VP and BVP methods.

4.3. Evaluation of BVP Effectiveness

We evaluated the effectiveness of BVP for 1,880 dies from a single wafer. The initial die locations used for BVP were randomly selected. Based on the prediction error from the initial samples, BVP iteratively determines the next best sampling location. In our experiment, we sampled 700 dies initially and iteratively chose additional samples. For the rest of the dies that were not sampled, their performance values were predicted and the corresponding prediction errors were calculated. We evaluated the prediction accuracy for sample sizes ranging from 800 to 1500.

Figure 7 compares the prediction errors of BVP and VP for test items #14 (a receiver current) and #33 (a power measurement). All the sampling locations for VP are randomly selected. The x-axis shows the sample size, and y-axis is the prediction error. For test item #14, a predictable item, BVP reduces the prediction error by up to 39% in comparison with VP using the same sample size. For test item #33, BVP reduces the prediction error by up to 21%. When the sample size becomes relatively large (e.g., 1,500 samples in this experiment), VP and BVP achieve similar prediction accuracy.

These results demonstrate the importance of selecting good sampling locations. For predictable items, BVP helps achieve the same prediction accuracy with a significantly smaller sample size (i.e., fewer measurements required). On

the other hand, because the spatial variations of highly-predictable items are strongly correlated, random sampling used by VP is sufficiently good and BVP has little additional benefit. For both predictable and highly predictable items, we can employ BVP during the pre-test analysis to fully characterize one wafer and then determine the sampling locations that best capture spatial correlations with minimum prediction error.

5. Conclusion

In this paper, we explore the application of VP to replace a large number of test measurement by prediction for mixed-signal/RF circuits. We first analyze the predictability of each test item using an iterative sampling-and-validation approach and determine the sample size for each wafer. A case study on an industrial RF transceiver demonstrates that more than 75% of the test items can be predicted using VP and the estimated speedup of overall test time is about 2.36x.

Acknowledgments

This work is partially supported by the Gigascale Systems Research Center (GSRC) and the Center for Circuits and System Solutions (C2S2), two of six research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation program. This work is also supported in part by the National Science Foundation under contract CCF–0915912.

References

[1] K.-T. Cheng and H.-M. Chang, "Recent Advances in Analog, Mixed-Signal, and RF Testing," *IPSJ Transactions on System LSI Design Methodology (TSLDM)*, vol. 3, pp. 19-46, Feb. 2010.

[2] K. Arabi, "Mixed-signal test impact to SoC commercialization," in *Proc. of 28th VLSI Test Symposium* (VTS), 2010.

[3] H.-G. D. Stratigopoulos and Y. Makris, "Error moderation in low-cost machine learning-based analog/RF testing," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 2, pp. 339–351, 2008.

[4] H.-G. D. Stratigopoulos, P. Drineas, M. Slamani, and Y. Makris,
"Non-RF to RF test correlation using learning machines: A case study," in *Proc. of 25th VLSI Test Symposium* (VTS), 2007.

[5] P. N. Variyam, S. Cherubal, and A. Chatterjee, "Prediction of analog performance parameters using fast transient testing," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 21, no. 3, pp. 349–361, 2002.

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

[7] 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," in *Proc. of IEEE/ACM International Conference on Computer-Aided Design* (ICCAD), pp. 47-54, 2010.

[8] W. Zhang, X. Li and R. Rutenbar, "Bayesian virtual probe: minimizing variation characterization cost for nanoscale IC technologies via Bayesian inference," in *Proc. of IEEE/ACM Design Automation Conference* (DAC), pp. 262-267, 2010.

[9] X. Li, R. Rutenbar, and R. Blanton, "Virtual probe: a statistically optimal framework for minimum-cost silicon characterization of nanoscale integrated circuits," in *Proc. of IEEE/ACM International Conference on Computer-Aided Design* (ICCAD), pp. 433-440, 2009.

[10] S. Sunter and A. Roy, "BIST for phase-locked loops in digital applications," in *Proc. of International Test Conference* (ITC), 1999.

[11] 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," in *Proc. of International Test Conference* (ITC), 2010.

[12] Y. Xing, "Defect-Oriented Testing of Mixed-Signal ICs: Some Industrial Experience," in *Proc. of International Test Conference* (ITC), 1998.