

Integrated Optimization of Semiconductor Manufacturing: A Machine Learning Approach

Nathan Kupp* and Yiorgos Makris†

*Department of Electrical Engineering, Yale University, New Haven, CT 06511

†Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX 75080

Abstract—As semiconductor process nodes continue to shrink, the cost and complexity of manufacturing has dramatically risen. This manufacturing process also generates an immense amount of data, from raw silicon to final packaged product. The centralized collection of this data in industry information warehouses presents a promising and heretofore untapped opportunity for integrated analysis. With a machine learning-based methodology, latent correlations in the joint process-test space could be identified, enabling dramatic cost reductions throughout the manufacturing process. To realize such a solution, this work addresses three distinct problems within semiconductor manufacturing: (1) Reduce test cost for analog and RF devices, as testing can account for up to 50% of the overall production cost of an IC; (2) Develop algorithms for post-production performance calibration, enabling higher yields and optimal power-performance; and, (3) Develop algorithms for spatial modeling of sparsely sampled wafer test parameters. Herein these problems are addressed via the introduction of a model-view-controller (MVC) architecture, designed to support the application of machine learning methods to problems in semiconductor manufacturing. Results are demonstrated on a variety of semiconductor manufacturing data from TI and IBM.

I. INTRODUCTION

The advent of the modern integrated circuit has created a immense market for semiconductor devices, surpassing the \$300 billion mark in 2011. Semiconductors are wholly pervasive in today's world, and are integrated into a wide spectrum of products. These products present dramatically different design constraints. For consumer electronics, low cost is the key driver; despite high manufacturing volume, profit margins are typically small. On the other hand, automotive and defense applications demand high reliability and security, with considerably lower manufacturing volume. Although increasingly sensitive to power consumption, traditional server and desktop computing applications remain primarily performance driven. Finally, the emergence of mobile devices has created a demand for semiconductor products with extremely low power consumption. Manufacturing such products that meet all of these constraints has proven to be extremely challenging.

Manufacturing semiconductor devices that meet the performance targets of such a broad range of applications is very challenging. However, throughout the manufacturing and test process, a tremendous amount of data is generated and collected, from raw silicon through to final packaged product. All of this data is collected and stored in semiconductor manufacturing information warehouses as part of the manufacturing process, and it is used by many disparate engineering teams during the manufacturing flow. The availability of this data is laden with opportunities for improving the manufacturing flow with statistical learning methods.

To date, the majority of statistical data analysis taking place within semiconductor manufacturing has been stratified throughout the manufacturing and test process, and limited to within-group process or test team work. Process data and inline measurements are typically employed to monitor the process, drive yield learning, and track manufacturing excursions. Beyond identifying device pass/fail status, test data is often used to search for yield detractors, albeit in a rather ad hoc fashion. As process variability dramatically increases with smaller process nodes, studying data jointly from both the manufacturing and test flow is becoming increasingly important, as isolated statistical data analysis is likely to miss important intra-process/test correlations.

In this work, a novel, integrated approach to semiconductor data analysis is introduced, constructed as a model-view-controller (MVC) framework. This design pattern, originally described in [1] as a framework for user interfaces, is adapted herein to semiconductor manufacturing and statistical data analysis. The proposed approach synthesizes semiconductor manufacturing practices with the latest advances in statistical learning theory and modern software engineering disciplines into a unified framework for solving data analysis problems within semiconductor manufacturing and test. Herein, several such problems are addressed, as case studies in the application of the proposed framework.

The remainder of this paper is organized as follows. In Section II, the model-view-controller framework is introduced and discussed. In Section III, a set of statistical algorithms addressing problems in low-cost testing of analog and RF devices is discussed. In Section IV, algorithms for the post-production calibration of analog and RF device performances are introduced. Section V addresses spatial modeling of wafer parameters via Gaussian process models. Lastly, Section VI presents some conclusions.

II. MODEL-VIEW-CONTROLLER FRAMEWORK

In this work, a framework for applying modern machine learning methods to semiconductor manufacturing is introduced. The proposed solution addresses the data stratification issues of semiconductor manufacturing by introducing a set of useful abstractions for analyzing and constructing integrated statistical models on semiconductor data. Moreover, the proposed approach incorporates state-of-the-art software engineering methods and statistical learning theory into a unified framework. To do this, we adopt the software engineering design pattern known as model-view-controller (MVC) [1]. MVC is a design pattern that is already widely used throughout

software engineering, albeit traditionally for architecting user interfaces. Herein, we demonstrate that MVC is also a suitable approach for solving machine learning problems in semiconductor data analysis and detail the various components of a proposed MVC architecture. We also describe an added fourth layer of the proposed MVC framework, entitled *Recipes*. Recipes, in the context of the proposed MVC framework, are simply flows that describe compositions of Model, View, and Controller components into solutions for semiconductor data analysis.

1) *Models*: The “model” component of the proposed MVC architecture describes a data representation constructed on top of relational databases or flat data files. These models can effectively be considered as high-level abstractions that simplify data flows, while avoiding issues of database infrastructure, storage, and integrity. For example, a chip model may have $\{x, y\}$ coordinates, measurement data, and a wafer ID. A wafer model then would be treated as a collection of chip models, along with any inline parametric measurements. Thus, models encapsulate key data segments from throughout the entire semiconductor fabrication process, and form the atomic elements that are transported and analyzed throughout the MVC architecture.

2) *Views*: The “view” component of the MVC architecture provides a means for the process or test engineer to directly observe and interact with semiconductor data analyses during execution. Views are comprised of graph, chart, or table-generating code, collectively entitled presentation logic. In the context of semiconductor data analysis, such views are designed to provide the engineers with information on the algorithms in use and present related statistical data. The important feature of views is the complete separation of presentation logic from the statistical algorithms deployed by the system.

3) *Controllers*: The “controller” component of the MVC framework performs atomic actions on data encapsulated in the models. For semiconductor data analysis, the controllers we have implemented as statistical algorithms, including both supervised and unsupervised learning methods: linear regression, linear discriminant analysis, k-Nearest Neighbors, support vector machines (SVMs), etc. The controller implementations are designed with simplicity in mind: the user may simply pass a collection of chip objects to a controller, and the controller will seamlessly construct appropriate statistical models with sensible parameter choices. For many applications, some fine tuning is beneficial, and the controller implementations provide complete parameterizations of the underlying statistical algorithms to permit the user to modify all of the free parameters. By implementing controllers in this fashion, the complexity of statistical algorithms is only exposed when necessary.

4) *Recipes*: The final component of the proposed MVC framework is a “recipe”. Within the framework, a recipe is a composition of models, views, and controller elements into a solution for a particular problem. The core framework has been built with a variety of recipes targeted at various

semiconductor data analysis problems. For example, an ATE simulator recipe has been developed and included with the MVC framework, which provides a chip iterator that regularly emits chip objects from device data files.

In Figure 1, we display an overview of the framework as applied to semiconductor manufacturing; the proposed framework overlays seamlessly on top of the existing semiconductor manufacturing flow. In the following sections, the versatility of the proposed framework is demonstrated through its application to problems in low cost testing, performance calibration, and wafer spatial modeling.

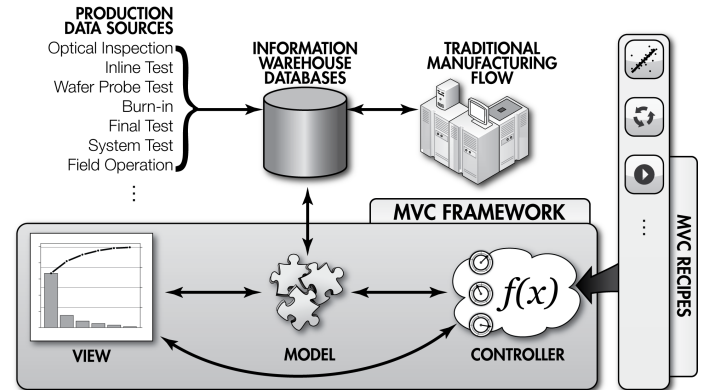


Fig. 1. MVC Framework: Overview

III. LOW-COST TESTING FOR ANALOG AND RF DEVICES

During semiconductor fabrication, every fabricated device must be explicitly tested in order to guarantee that it meets the original design specifications. For analog and RF devices, the test process involves explicitly measuring specification performances for every device. Such testing identifies latent defects that are due to the various sources of imperfection and variation in the fabrication process. Defects can present as either catastrophic or parametric. Catastrophic defects typically lead to a complete malfunction of the device and are consequently easily detected by inexpensive tests. Parametric faults, which are caused by excessive process variations, are considerably more difficult to detect.

To catch such parametric faults, RF circuits are typically tested directly against the parametric specification performances. Although this approach is highly accurate, it comes at a very high cost, which can amount up to 50% of the overall production cost. Given that RF circuits typically occupy less than 5% of the die area, there is great industrial interest in the reduction of RF test cost [2], [3], [4]. The high cost of RF test is due to the expense of automated test equipment that is required, and, on the other hand, the lengthy test times that result from a sequential measurement approach. In response, a variety of low-cost test methods have been introduced [5], [6], [7], [8], [9]. In this section, various problems within the low-cost testing space are addressed.

A. Non-RF to RF correlation-based specification test compaction

One way to address the high test cost for RF devices is to remove all RF tests from the test set, and employ statistical models to predict the untested measurement outcomes. The framework of this non-RF to RF correlation-based specification test compaction process is depicted in Fig. 2. The learning phase relies on a training set of devices, on which both the non-RF performances and the RF performances are explicitly measured. Based on this information, statistical correlation models are learned, predicting each excluded performance as a function of the non-RF performances of a device, or a subset of those performances. Subsequently, for every new device in production, only the selected non-RF performances are explicitly measured, while the untested non-RF and RF performances are predicted through the learned correlation models. A pass/fail decision is made by comparing the explicitly measured non-RF performances and the predicted performances to their specifications. Thus, an RF ATE is needed only for characterizing the small number of devices in the training set but is not necessary during production testing.

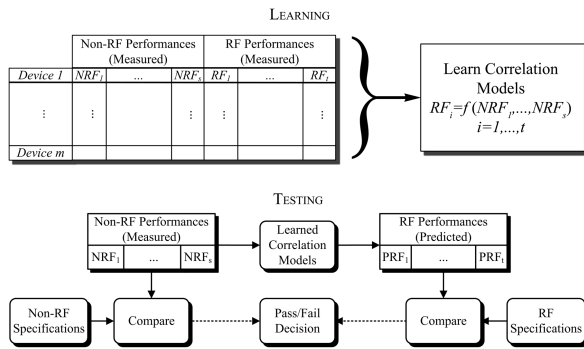


Fig. 2. Non-RF to RF Correlation-Based Testing

However, this test compaction procedure is often subject to prediction error beyond industrial acceptable ranges of test escapes. Consequently, providing a confidence level indication along with the predicted performances could go a long way. Devices for which this confidence level is low can then be identified and discarded or retested, as shown in the two-tier test approach of Fig. 3.

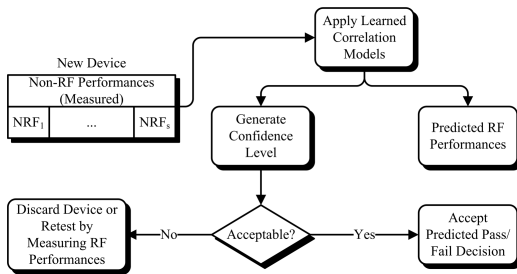


Fig. 3. Retesting when prediction confidence is low

1) *Confidence estimation*: In this work, we introduce a method termed *confidence estimation* for deciding whether pass/fail predictions are sufficiently accurate. It employs an

additional learning phase, wherein a Support Vector Machine (SVM) [10] is trained to separate the non-RF measurements into regions that are trusted or untrusted with regards to the pass/fail decisions of the correlation models.

To do this, correlation models are initially learned from training set devices. Then, the learned correlation models are used to make pass/fail predictions on a hold-out set, and the devices in the hold-out set are relabeled as correctly or incorrectly predicted. In the second step, an SVM is trained to learn the boundary partitioning the predicted performance space into two subspaces: the area wherein correct predictions occurred (trusted), and the area wherein incorrect predictions occurred (untrusted). Lastly, the utilization of the SVM during the testing phase (for a toy two-dimensional example with two non-RF features) is conceptually demonstrated in Figure 4. The pass/fail prediction of the correlation models is accepted only for devices that the SVM classifies as trusted, while the rest of the devices are retested.

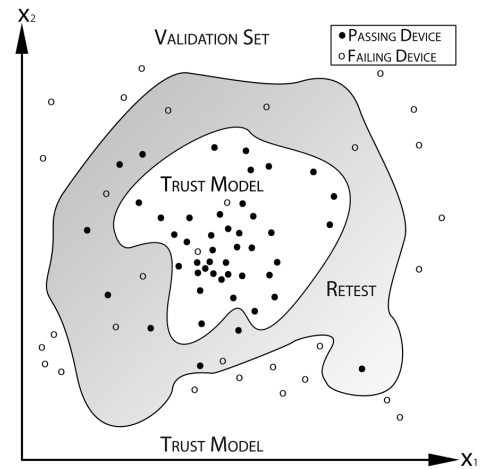


Fig. 4. Confidence estimation - Testing phase

2) *Experimental results*: In order to assess the relative effectiveness of confidence estimation, we use production test data from a zero-IF down converter for cell-phone applications, designed in RFCMOS technology and fabricated at IBM. The device is characterized by 143 performances, 72 of which are non-RF (i.e. digital, DC, low frequency) and 71 are RF. The test dataset includes performance measurements for 4450 devices across 3 lots. Of these devices, 4141 pass all the specification tests while 309 fail one or more specification tests. The passing and failing devices are each randomly split into three subsets of equal size, and used as the training set, the hold-out set and the validation set, respectively. The results for the proposed confidence estimation method are shown in Figure 5. As can be observed, both the number of retested devices and the test error of the confidence estimation method are reduced as compared to existing prior art methods of defect filtering [11], [12] and guard banding [13] methods.

3) *Summary*: Specification test compaction through non-RF to RF performance correlation promises significant test cost reduction. Yet, in order to meet industry-level DPM

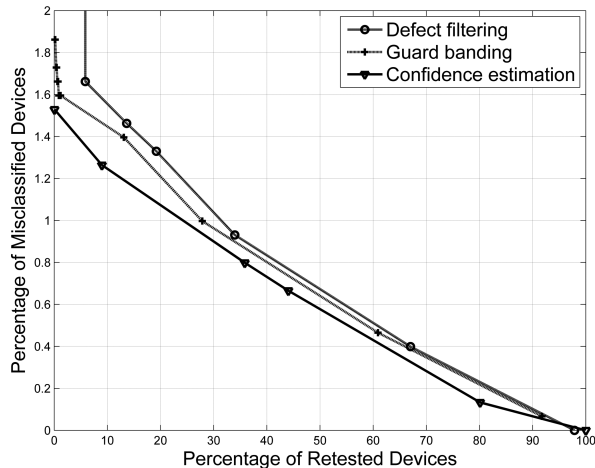


Fig. 5. Comparative results

standards, such compaction relies on efficient methods for boosting the accuracy of the correlation models and exploring the trade-off between the test error and the number of devices that need to be retested through complete specification testing. As demonstrated experimentally using production test data from a zero-IF down-converter fabricated by IBM, the proposed method facilitates an efficient exploration of the tradeoff between test cost and test accuracy for non-RF to RF test compaction, even in the region of very low DPM levels.

B. On Proving the efficiency of low-cost RF tests

Despite the number of alternative or low-cost RF test approaches which have been proposed to date, the industry is largely reluctant to replace traditional RF testing. The primary reason is the lack of automated tools for evaluating a new test approach fast and early at the design and test development phases, before moving to production test. It may be easy to estimate the area overhead incurred by a built-in test solution and to study to what degree it degrades the device performances, yet it is extremely difficult to estimate the incurred indirect costs, that is, the resulting test errors. This section presents a new method and case study of test metrics estimation. Specifically, the aim is to prove the equivalence of low-cost On-chip RF Built-in Tests (ORBiTs) [4] to the traditional RF specification tests, based solely on a small data set obtained at the onset of production. Our results are validated on a much larger data set containing more than 1 million Bluetooth/Wireless LAN devices fabricated by Texas Instruments.

A straightforward way to characterize the low-cost test system under consideration and obtain accurate, parts-per-million (ppm) test metric estimates is to take a very large set of fabricated devices, say 1 million, and apply the test system to each device, recording test metrics on each. However, given the extremely high cost, this is not a sustainable practice for evaluating candidate low-cost test systems. In this section, we employ a general technique for obtaining ppm-accurate

test metric estimates, originally developed in [14], and we examine for the first time its potential on a real-world case study. This technique is able to elegantly achieve the objective of providing such accurate estimates, without the extreme cost associated with having to consider millions of fabricated devices. It is based on the statistical methodology of non-parametric kernel density estimation (NKDE), as shown in Figure 6. The underlying idea is to rely on a small set of representative devices to estimate the joint non-parametric probability density function of specified performances and low-cost tests. Thereafter, the estimated density is sampled to generate a large synthetic set of device instances from which one can readily compute test metrics using relative frequencies. Moreover, we are able to provide a case study demonstrating equivalence of our proposed system of Figure 6 and the true ppm metrics obtained via explicitly testing 1 million devices.

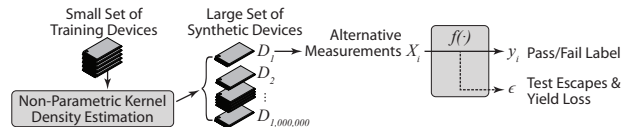


Fig. 6. Low-cost method for obtaining parts-per-million test metric estimates

The ORBiTs have been proven to be generally very efficient in such replacements, but this knowledge was acquired only after measuring millions of RF device instances with the dedicated built-in test circuitry. In this section, we try to answer the following question: Is it possible to estimate values of the test metrics close to true ones while employing in the analysis a small set of RF devices that we obtain at the onset of production? In this case, we will be able to decide on the efficiency of the ORBiTs early in the process without having to wait for a large volume of silicon data to reach a safe conclusion. This type of proactive analysis is very important in cases where the low-cost tests are found later on to be inefficient. It allows to convince test engineers about the efficiency of an approach, to identify shortcomings and come up with remedies for refining an approach, or abandon an approach altogether if it is deemed not to be equivalent to the standard specification test approach.

1) *Test metrics estimation method:* It is in this context that we introduce a novel methodology originally proposed in [14] to obtain test metric estimates with ppm accuracy, while side-stepping the cost associated with exhaustively testing millions of devices. This technique, based on NKDE, permits dramatically enriching the validation set with synthetic device instances reflective of the true device population. With this large synthetic device set in hand, we are able to produce test metric estimates using relative frequencies. NKDE relies on a small Monte Carlo run (e.g. on the order of a few thousands devices) to generate a synthetic device sample with population statistics nearly identical to the 10^6 -order population. The underlying idea is to estimate the joint probability density function of ORBiTs and specification tests based on the

small Monte Carlo run. The NKDE approach makes no *a priori* assumptions about the parametric form of the generative probability density function (e.g. Gaussian) and allows the available simulation data to speak for themselves.

2) *Feature selection*: Often when dealing with early characterization data sets, a large number of measurements are available which are later pruned for the final test set. This provides a wealth of data, but also presents a case of the well-known “curse of dimensionality”, the law that by adding dimensions, one exponentially increases data sparsity. This data sparsity can cause learning algorithms to have high-variance classification boundaries and poor generalization capability. A key component of our analysis was to perform feature selection on the very high dimensional space of available ORBiTs, projecting device test signatures into a lower dimensional subspace. Herein we employ a supervised feature selection method known as Laplacian score feature selection (LSFS) [15] to rank ORBiTs and subsequently reduce dimensionality.

3) *Experimental results*: To confirm the efficiency of our approach in providing early estimates of test metrics, we employed a Texas Instruments data set from a total of more than 1.1 million devices. The devices are collected from 176 wafers and each wafer has between 6,000 and 7,000 devices. For each device, the data set contains the ORBiT measurements and the specified performances in the data sheet. Specifically, there are 739 ORBiT measurements and 367 performances. Some ORBiT measurements and performances are discrete-valued. The test metrics estimation method discussed in Section III-B1 is defined only for continuous variables. Therefore, in our analysis we considered only the continuous ORBiT measurements and performances, which number 249 and 264, respectively.

In our analysis, we focused on replacing the single most sensitive specification test, that is, the test that corresponds to most commonly failing performance across all wafers. We denote this performance by P . To predict the test metrics, we only use devices from the first wafer and we employ the NKDE technique to generate 1 million additional synthetic devices, in order to achieve ppm levels of accuracy, as illustrated in Figure 6. We emphasize that our use of a single wafer is purely to demonstrate the efficacy of the method in extremely challenging circumstances; in reality this sample may include an arbitrarily large training sample.

4) *Removing Outliers*: From the first training wafer, we remove outliers via a “defect filter”, for two reasons. First, we do not wish outliers with non-statistical signatures to have leverage over the feature selection process; the retained features should excel at discerning the more difficult parametric fails rather than the relatively easy-to-detect catastrophic fails. Second, the test metrics estimation method itself relies on estimating a probability density function, thus we should avoid using outliers for this purpose since they are non-statistical in nature and are not generated by the same probability distribution which assumes only process variations. This procedure results in the removal of approximately 3% of the device instances from the first wafer training set. Note that this step

is necessary only for the training set, and subsequent outlier fails are not removed in this fashion.

5) *Feature Selection: Reducing the dimensionality of ORBiTs*: Fitting a classifier boundary in a sparse, high dimensional space can be error-prone due to the consequent variance of the fitted class boundary. For this purpose, we employ LSFS to reduce the dimensionality of the problem. In particular, for each of the 249 ORBiTs, we compute and rank the Laplacian scores. In this experiment, we retained 7 ORBiT features. It should be noted that a lower dimensionality also maximizes the efficacy of NKDE, as it is also vulnerable to the “curse of dimensionality”.

6) *Information-rich training set*: It turns out that even in relatively densely-populated spaces, classifier performance can benefit by further increases in data density. Specifically, it is not advisable to attempt to directly fit a classification boundary to a severely unbalanced population, as the classifier tends to always label subsequent instances as the dominant class after training. To combat this effect and improve classifier performance by increasing data density in the training set, we also employed non-parametric density estimation to generate synthetic training instances. To do this, we fit the joint probability density function of the instances from the first wafer. We sample the empirical probability density function to generate a much larger, information-rich training set that has a more balanced population of good, faulty, and critical devices across the decision boundary in a similar fashion to the approach taken in [16].

7) *Summary and Results*: Assembling the preceding steps, we arrive at the complete analysis approach shown in Figure 7. The training set is employed to train the SVM classifier to assign limits on the 7 ORBiTs in the form of a hyper-surface boundary. The limits are used to obtain the ground truth test escape and yield loss values for each wafer, denoted by T_E and Y_L , respectively. These values are averaged to obtain the ground truth ppm test escape and yield loss measured over the complete device population in hand, denoted by \bar{T}_E and \bar{Y}_L , respectively. The same limits are used on the synthetic device set generated from the first wafer, in order to obtain early ppm estimates of the test escape yield loss, denoted by \hat{T}_E and \hat{Y}_L , respectively.

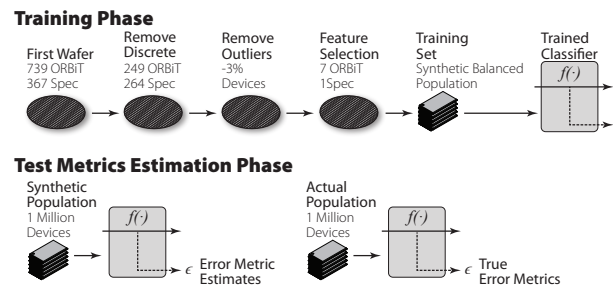


Fig. 7. Summary of experimental approach

The results are shown in Figure 8. As can be observed, test escape is slightly underestimated, and yield loss is very slightly

overestimated. Specifically, the true values are $\bar{T}_E = 0.7286\%$ and $\bar{Y}_L = 4.387\%$, whereas the early estimates are $\hat{T}_E = 0.4302\%$ and $\hat{Y}_L = 4.401\%$, that is, a difference of $\Delta T_E = 0.2984\%$ and $\Delta Y_L = -0.014\%$. We remind that the objective of the section is not to propose an SVM-based low-cost test technique, but to evaluate a candidate low-cost test technique at an early phase. Moreover, we evaluated the scenario where a subset of ORBiTs replaces the most sensitive specification test, and not the general case where the complete suite of ORBiTs is used to replace irrespectively all specification tests.

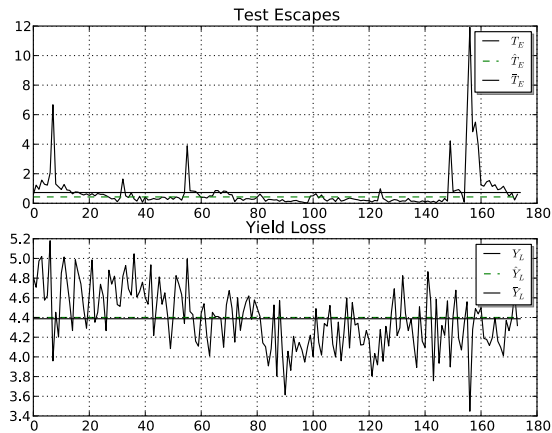


Fig. 8. Prediction across all wafers

C. Summary

In this section we presented a method for providing accurate, parts-per-million estimates of test metrics without incurring the cost associated with simulating or testing millions of devices. A comparatively small set of RF devices from a single wafer tested at the onset of production coupled with the proposed NKDE-based sampling are used to generate one million synthetic device samples, on which we are able to evaluate test escape and yield loss test metric estimates. Furthermore, we have demonstrated our test metric estimates to be very close to the true values measured on more than one million devices from Texas Instruments.

IV. POST-PRODUCTION PERFORMANCE CALIBRATION

In modern analog device fabrication, circuits are typically designed conservatively to ensure high yield. Otherwise, yields may be low due to process variation driving devices beyond specification limits. Thus, the analog designer often finds himself doubly constrained by performance and yield concerns. However, the demand for high performance, high yield analog devices is relentless. As such, recent interest has been shown in producing analog devices that are tunable after fabrication by introducing “knobs” (post-production tunable components) into the circuit design. By adjusting the knobs, some devices that would simply be discarded under the traditional analog test regime can be tuned to meet specification limits and thereby function correctly. These tunable devices would permit analog designers to create aggressively high-performance

integrated circuits (ICs) with expectations of reasonable yield. Alternatively, conservative designs could be produced with nearly-perfect yields.

To date, post-production performance calibration has not achieved widespread use due to the perceived complexity and cost of implementation. This is not an unreasonable perception: knobs have apparently complex interdependent effects on performances, and iterative specification test-tune cycles to explore the large space of knob settings are prohibitively costly. In this work, we outline several key observations which appropriately constrain the free parameters of performance calibration methodologies to enable straightforward cost-effective implementation. Moreover, we develop and present a cost model which permits direct comparison of performance calibration to specification test and other state-of-the-art practices.

Implementation of performance calibration requires selection of key parameters of the circuit (voltage, capacitance, etc.) as knobs. Additional circuit elements are added to enable post-production modulation and on-chip storage of these parameters. By setting knob values, the performances of the circuit can be dynamically modified and improved to meet specification limits. After the knobs are in place, a method for circuit tuning must be devised. Herein we focus specifically on addressing this problem. A first-order solution is to simply iteratively test across all knob settings until a setting that results in a passing device is found. Knob setting selection can also be performed by attempting to find some optimum for the circuit, i.e. searching for the lowest possible power setting.

A. Midpoint low-cost test performance calibration

Clearly, these approaches are not economically feasible given the high cost of specification test. Herein lies the benefit of adopting alternate test as a basis for performance calibration, as alternate tests can be an order of magnitude less costly to perform. In this work we introduce a novel alternate test-based performance calibration method, entitled midpoint alternate test-based performance calibration. To manage the cost of performance calibration, we must modify the exhaustive test approach to reduce the large number of measurements (alternate test or otherwise) which must be collected. We do this by making an important observation: knob variation and process variation *orthogonally* act on device performances. Thus, we can separately model each axis of variation and construct a composite model which accounts for both.

Alternate tests are designed to correlate well with device performances. Implicitly, this means that we can already model process variation from the alternate tests. To model knob variation, we examine a simulated process-variation free device. By modeling knob effects in simulation, we only need to explicitly measure alternate tests at a *single* knob setting in production test, where all knobs are set to their respective midpoint values. This set of midpoint alternate test measurements can be used to predict performances at *all* of the knob settings, as shown in Figure 9.

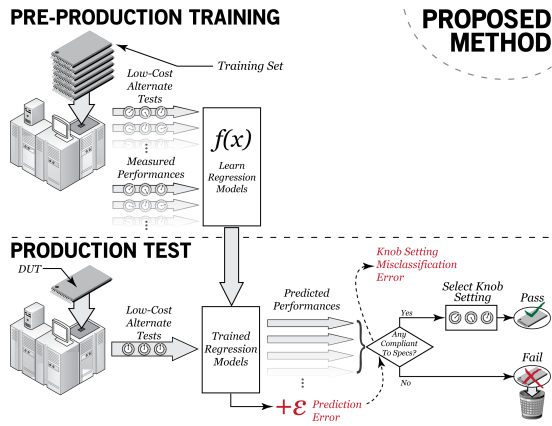


Fig. 9. Midpoint low-cost test performance calibration

B. Knob setting selection

Once we have modeled both knob effects and process variation effects, the models must be employed to inform knob setting selection decisions on each device in the test set, where limited information about the device is available to us. In our work, we have chosen to predict performances for each knob setting for every device. This allows us to accomplish two things. First, we partition the test set into unhealable and healable regions by determining for every device in the test set whether at least one knob setting is available which will heal it. Second, for every healable device in the test set we predict a family of knob settings which will heal it. In this work, we use both Mahalanobis distance from specification planes and predicted power as optimizing metrics to select the best knob setting from this family of predicted-to-heal knob settings.

C. Experimental validation

To validate our methods, we designed a cascode low-noise amplifier (LNA) in $0.18\mu\text{m}$ CMOS. In this section, we document our design choices and show experimental results for the proposed midpoint alternate test-based performance calibration method. We selected the RF LNA as our platform for experimental validation, as it is one of the most frequently used components in commercial transceiver RFICs. Among the numerous possible LNA architectures, we chose one of the most widely-adopted designs, the cascode topology. To perform post-production performance calibration, we modified the LNA topology to include tunable circuit elements. In our device design, we selected three key bias voltages to include as tuning knobs, as these provided maximal control over performances.

Along with the LNA, we designed and implemented an on-chip amplitude sensor and on-chip signal generator, for collecting alternate test data. With an appropriate choice of input signals, the alternate test measurements produced by the amplitude sensor/signal generator pair have been demonstrated to be well-correlated with LNA performances.

The layout-level LNA was used to collect performance data across all knob settings of each device. For the alternate test data, two amplitude detectors were added at the input and

output of the LNA, and both were measured with stimuli provided by the RF signal generator. Two different frequencies of the RF signal generator were employed, for a total of 4 alternate test measurements collected per knob setting per device.

1) *Dataset*: For our experiments, we created 1,000 instances of the LNA with process variation effects included to simulate a production environment. The 3 knobs in the LNA designed for our experiment were assigned 3 discrete settings (i.e., 1.6V, 1.8V, 2.0V) for a total of $3^3 = 27$ possible knob positions. On every device in our dataset, we collected four performances: S11, Noise Figure (NF), Gain, and S22. We also collected a power measurement and the four low-cost amplitude sensor (peak detector) alternate test measurements. Thus, for every device there are 9 figures of merit, such that the entire dataset is a $1,000 \times 27 \times 9$ matrix. In production it is infeasible to measure all circuit performances on every device (elements of this matrix) are explicitly measured.

As stated previously, if we are to model the circuit response to knob and process variation, an initial training set must be generated which includes the relationships we wish to model. For example, if we wish to predict circuit performances at every knob setting, these performances must be explicitly measured for a small training set in order to construct our models. Once these models are constructed, they can be used to predict circuit performances for the remaining circuits. For the experiments which required training statistical models, we split the dataset 50/50, training on data from 500 devices and predicting on the remaining 500. We also performed 10 cross-validations to ensure statistical stability of the reported results.

2) *Performance calibration: midpoint alternate test*: Using the proposed performance calibration methodology, we classified devices as healable or unhealable, with a success rate demonstrated in the confusion matrix of Table I. Thus, due to the use of alternate test, an approximately 0.62% test escape rate and a 0.48% yield loss rate are introduced, for a slightly over 1% total error rate.

		Actual	
		Unhealable	Healable
Predicted	Unhealable	1.98%	0.48%
	Healable	0.62%	96.92%

TABLE I
MIDPOINT ALTERNATE TEST

a) *Knob Setting Selection*: As we outlined in Section IV-B, once the healable devices have been identified using midpoint alternate test, knob setting selection is performed by employing the Mahalanobis distance or the predicted power knob setting selection metric. Presented in Figure 10 is the power vs. correct-heal tradeoff for the knob setting selection optimality metrics: minimum power, median power, and maximum distance. As can be seen from the figure, the Mahalanobis distance metric achieves a near-perfect 99.2% correct-heal rate, at the expense of high power consumption, whereas minimizing power (as expected) substantially improves power consumption, while slightly increasing error.

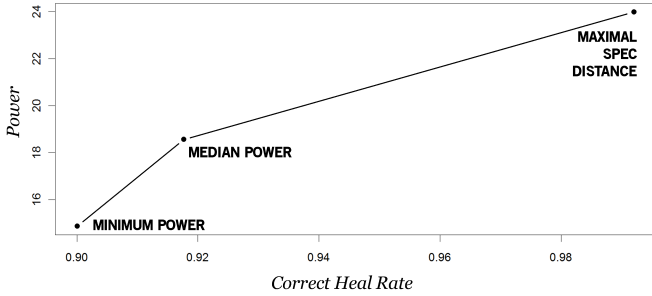


Fig. 10. Power-Prediction Quality Tradeoff

D. Summary

We have demonstrated that appropriate modeling of knob and process variation enables highly successful performance calibration. The proposed midpoint alternate test is a cost-effective means of introducing performance calibration methodologies into an analog/RF device test flow. Indeed, it overcomes the limitations of both iterative approaches and two-model approaches by implementing a *single model* requiring a *single alternate test measurement* step to perform tuning. This method achieves highly accurate healable/unhealable classification, with a 0.62% test escape rate and a 0.48% yield loss rate, and a 99.2% correct-heal rate using the distance metric to select a knob setting on the healable devices. Finally, we have demonstrated experimentally that we can decouple the training set size from the number of knob settings N_K , requiring only a small random sample of alternate tests and performances from a handful of devices to sufficiently learn the statistics of knob and process variation.

V. WAFER SPATIAL MODELING

In the course of semiconductor manufacturing, various wafer-level measurements are collected throughout the manufacturing and test process. In this section, a statistical algorithm that operates on sparsely sampled sets of measurements is introduced. The proposed approach creates interpolative models, enabling prediction of parameters spatially across a wafer. While there are many use cases for spatial interpolation of semiconductor manufacturing data, this section explores two particularly important applications.

First, statistical interpolation for e-test measurements is presented. During manufacturing, e-test measurements (also known as inline or kerf measurements) are collected to monitor the health-of-line and to make wafer scrap decisions preceding final test. These measurements are typically sampled sparsely across the surface of the wafer from between-die scribe line sites and include a variety of measurements that characterize the wafer’s position in the process distribution. The proposed methodology permits process and test engineers to examine e-test measurement outcomes at the die level, and makes no assumptions about wafer-to-wafer similarity or stationarity of process statistics over time.

Second, we present a statistical interpolation approach for RF tests. Given the high cost of RF automated test equipment (ATE) and lengthy test times, the incurred test cost per device can be quite high. Various statistical methodologies have been proposed to address this problem by attempting to reduce the number of RF tests required (test compaction), introduce new alternative tests [5], or build machine learning models to learn classification boundaries separating passing and failing populations of devices [9].

In this work, we introduce a Gaussian process model-based probe test prediction method [17]. Instead of completely eliminating all RF tests, we collect them on a small sample of devices on each wafer. The probe test outcomes of these die are then used to train spatial regression models, and subsequently, these models are used to extrapolate probe test values for the remaining die on a given wafer. In most cases, this small sample is sufficient for us to extract wafer variation statistics for each probe test parameter and accurately model probe test outcomes at untested die locations.

A. Gaussian Process Models

Gaussian process models are birthed from the union of Bayesian statistics and the kernel theory of Support Vector Machines [10]. With Gaussian processes, we do not presume the generative function $f(\mathbf{x})$ is of linear form. Instead, we define a Gaussian process as a collection of random variables $f(\mathbf{x})$ indexed by coordinates \mathbf{x} , such that every finite set of function evaluations over the coordinates is jointly Gaussian-distributed. It can be shown [17] that this model form leads to regression models with a variety of useful properties.

For e-test parameter interpolation, our objective is to build per-wafer Gaussian process models that accurately estimate e-test parameter outcomes at previously unobserved die locations. By modeling variation on a per-wafer basis, we sidestep the need for the “median polishing” methodology of [18]. Our Gaussian process implementation was designed to record spatial prediction error across the surface wafer and provide a metric of prediction quality in terms of percent error.

For probe test measurements, we build per-wafer Gaussian process models of spatial variation by training on a small sample of explicitly tested devices and predicting all of the remaining test outcomes at unobserved wafer die locations. As with e-test interpolation, we capture the effectiveness of our proposed methodology by recording the percentage prediction error of our statistical model on each measurement and each wafer. To train the Gaussian process model for a particular wafer, we collect probe test data from a very small sample (approximately 20) die. These die are then used as a training set to train the Gaussian process model. The remaining die are collected as the test set on which we apply the trained model.

B. Experimental results: e-Test Interpolation

In this work, we demonstrate results on e-test data collected from industry HVM. Our dataset has in total 8,440 wafers, and each wafer has 269 e-test measurements collected from 21 sites sampled across the wafer. Leave-one-out cross validation

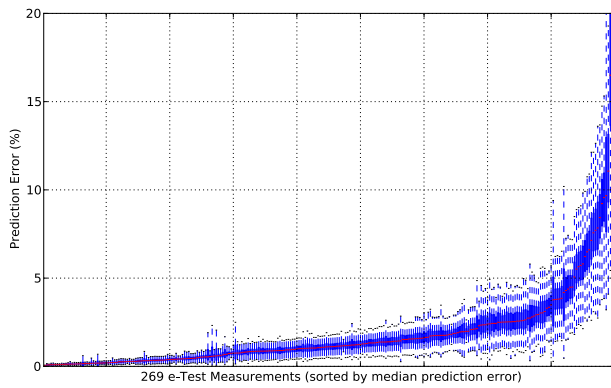


Fig. 11. Gaussian process prediction error for each e-test measurement

was used to characterize the prediction error at each of the 21 sites, and the mean cross-validation error across all 21 sites was collected for each e-test measurement and each wafer.

In Figure 11 we present the Gaussian process model prediction errors with 10%–90% error bars. The error ranges are quite small across the majority of e-test measurements, demonstrating that the variance of the errors is low over the entire set of 8,440 wafers. Importantly, this shows that the models are insensitive to process shifts over time, a result that is attributable to the fact that we train and deploy our models on a per-wafer basis. A tabular comparison of the proposed Gaussian process model approach versus Virtual Probe is provided in Table II, with overall mean error reported alongside mean training and prediction time per wafer across all e-test measurements. The timing measurements represent the mean total time required to construct and predict with all $269 \times 21 = 5,649$ models for a given wafer. Note that the proposed methodology consistently has lower error than Virtual Probe, while incurring an order of magnitude less runtime to construct and evaluate the predictive models.

Method	Overall Mean Percent Error	Avg. Running Time (per wafer)
Virtual Probe	2.68%	116.2s
Gaussian Process Model	2.09%	16.43s

TABLE II
VIRTUAL PROBE & GAUSSIAN PROCESS MODELS COMPARISON

C. Experimental results: probe test interpolation

Our probe test results are shown using probe test data from high-volume semiconductor manufacturing. The device under consideration is an RF transceiver with four radios. Our dataset has a total of 3,499 wafers with 57 probe test measurements collected on each device. Each wafer has approximately 2,000 devices, and a training sample of 20 devices were used on each wafer to train the spatial models. The models trained on these 20 devices were then used to predict the untested probe test outcomes at the remaining die coordinates.

In Figure 12, we present the Gaussian process model prediction errors with 10%–90% error bars. Note that the widths of the error bars are quite small, indicating that the prediction

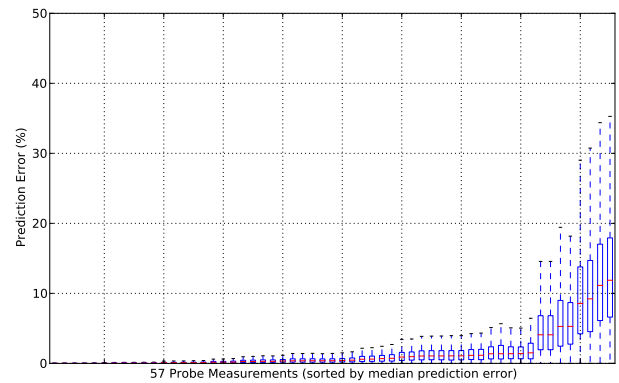


Fig. 12. Gaussian process prediction error for each probe test measurement

errors demonstrate low variance over the complete dataset of 3,499 wafers. Since we construct our statistical models on a per-wafer basis using a small sample from each wafer, the models are relatively insensitive to temporal process shifts.

The proposed methodology consistently outperforms Virtual Probe by an average of 16.5%, and in a few cases by more than 25%. In absolute terms, the overall mean prediction error of Virtual Probe across all probe test measurements and all wafers is 18.2%, while the overall mean prediction error for Gaussian process-based spatial models is only 1.71%, as shown in Table III. The per-wafer training and prediction time for Virtual Probe and Gaussian process models is also presented in Table III; the proposed methodology is extremely fast and requires less than a second to complete the full train-predict cycle for an entire wafer. The timing measurements represent the mean total time required to construct all 57 models and predict performances for all die on a given wafer. In summary, the proposed methodology consistently exhibits lower error than Virtual Probe, while requiring dramatically less runtime to construct and evaluate the predictive models.

Method	Overall Mean Percent Error	Avg. Running Time (per wafer)
Virtual Probe	18.2%	422.5s
Gaussian Process Model	1.71%	0.586s

TABLE III
COMPARISON OF VIRTUAL PROBE & GAUSSIAN PROCESS MODELS

D. Summary

In this section, a Gaussian process model-based methodology for generating spatial estimates of sparsely sampled e-test and probe parameters was presented. For e-test parameters, our Gaussian process model is able to generate extremely accurate predictions across more than 8,000 HVM wafers. For the majority of the parameters measured on these wafers, the Gaussian process model-based methodology demonstrates less than 4% error. Moreover, the distribution of prediction errors is tightly clustered across all of the wafers, indicating that our models are not affected by process shifts over time. Lastly, the proposed approach consistently outperforms Virtual Probe, on average by 0.5%, and in certain cases, by a significant

margin of almost 5%, while requiring an order of magnitude less runtime to evaluate on each wafer. Similar results are obtained for probe test parameter predictions; the proposed methodology enables dramatic reductions in probe test cost, by avoiding the requirement for dense application of costly probe tests. As demonstrated on more than 3,000 wafers, the proposed methodology requires only a very small sample (on the order of 1%) of die on each wafer to construct highly accurate spatial interpolation models. Despite this sparse sampling, a mean probe test prediction error of less than 2% is achieved, an order of magnitude lower than the existing state-of-the-art. Moreover, the proposed methodology is considerably faster to apply, requiring less than a second to train and predict on each wafer.

VI. CONCLUSION

In this work, an integrated solution for general data analysis and machine learning problems in semiconductor manufacturing was studied. The proposed MVC framework moderates the complexity of machine learning systems. This enables rapid iteration of candidate methods to find optimal solutions to challenging statistical problems involving large datasets, and forms a useful basis for solving general statistical problems involving semiconductor data.

To demonstrate the efficacy of the MVC framework as a solution for semiconductor data analysis, a set of challenging statistical problems in semiconductor manufacturing were posed and addressed. These problems spanned the breadth of semiconductor manufacturing, and incorporated both process and test data from a number of different circuit designs. Results were demonstrated on simulation data, as well as on production data consisting of millions of fabricated devices from two major semiconductor manufacturers: Texas Instruments and IBM. In particular, three distinct subspaces of semiconductor manufacturing data analysis were studied:

b) Low-Cost Testing: In this work, a set of solutions were proposed that dramatically increase the efficacy of low-cost test methods, by more optimally identifying confidence estimates and by generating parts-per-million accurate early estimates of test metrics for low-cost test systems.

c) Post-Production Performance Calibration: In this work, a single-test, single-tune algorithm for selecting optimal knob settings was proposed, enabling substantial yield improvements without requiring iterative test-tune-test cycles.

d) Wafer Spatial Modeling: In this work, a Gaussian process model-based predictive algorithm was proposed that enables highly accurate spatial predictions of wafer measurements at unobserved die locations. Moreover, the method was extended to enable test cost reduction: expensive tests can be sampled sparsely at only a few die locations per wafer, and the expensive test performances can be predicted at the unobserved die locations.

ACKNOWLEDGEMENTS

This research has been carried out with the support of the National Science Foundation (NSF CCF-1149463) and the

Semiconductor Research Corporation (SRC-1836.073). The author is supported by an IBM/GRC (Global Research Collaboration) graduate fellowship.

REFERENCES

- [1] Trygve M.H. Reenskaug, "Thing-model-view-editor – an example from a planning system," *Technical report, Institutt for informatikk, University of Oslo*, May.
- [2] J. Ferrario, R. Wolf, S. Moss, and M. Slamani, "A low-cost test solution for wireless phone RFICs," *IEEE Communications Magazine*, vol. 41, no. 9, pp. 82–88, 2003.
- [3] A. Abdennadher and S. A. Shaikh, "Practices in mixed-signal and RF IC testing," *IEEE Design & Test of Computers*, vol. 24, no. 4, pp. 332–339, 2007.
- [4] 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 *IEEE International Test Conference*, 2010, Paper 14.1.
- [5] P. N. Variyam and A. Chatterjee, "Specification-driven test generation for analog circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 19, no. 10, pp. 1189–1201, 2000.
- [6] S. Bhattacharya, A. Halder, G. Srinivasan, and A. Chatterjee, "Alternate testing of RF transceivers using optimized test stimulus for accurate prediction of system specifications," *Journal of Electronic Testing: Theory and Applications*, vol. 21, pp. 323–339, 2005.
- [7] S. Biswas, P. Li, R. D. Blanton, and L. Pileggi, "Specification test compaction for analog circuits and MEMS," in *Design, Automation and Test in Europe*, 2005, pp. 164–169.
- [8] S. Biswas and R.D. Blanton, "Reducing test execution cost of integrated, heterogeneous systems using continuous test data," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 30, no. 1, pp. 148–158, jan. 2011.
- [9] 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.
- [10] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.
- [11] J. B. Brockman and S. W. Director, "Predictive subset testing: Optimizing IC parametric performance testing for quality, cost, and yield," *IEEE Transactions on Semiconductor Manufacturing*, vol. 2, no. 3, pp. 104–113, 1989.
- [12] S. S. Akbay and A. Chatterjee, "Fault-based alternate test of RF components," in *Proc. of International Conference on Computer Design*, 2007, pp. 517–525.
- [13] R. Voorakaranam, A. Chatterjee, S. Cherubal, and D. Majernik, "Method for using an alternate performance test to reduce test time and improve manufacturing yield," Patent Application Publication #11/303,406, 2005.
- [14] H.-G. Stratigopoulos, S. Mir, and A. Bounceur, "Evaluation of analog/RF test measurements at the design stage," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 28, no. 4, pp. 582–590, Apr. 2009.
- [15] X. He, D. Cai, and P. Niyogi, *Laplacian Score for Feature Selection*, In NIPS. MIT Press, 2005.
- [16] H.-G. Stratigopoulos, S. Mir, and Y. Makris, "Enrichment of limited training sets in machine-learning-based analog/RF test," in *Design, Automation & Test in Europe*, 2009, pp. 1668–1673.
- [17] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.
- [18] F. Liu, "A general framework for spatial correlation modeling in VLSI design," in *Proc. of Design Automation Conference*, 2007, pp. 817–822.