Spatial Correlation Modeling For Probe Test Cost Reduction in RF Devices

Nathan Kupp*, Ke Huang†, John M Carulli Jr‡, 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
‡Texas Instruments Inc., 12500 TI Boulevard, MS 8741, Dallas, TX 75243

***Abstract*—Test cost reduction for RF devices has been an ongoing topic of interest to the semiconductor manufacturing industry. Automated test equipment designed to collect parametric measurements, particularly at high frequencies, can be very costly. Together with lengthy set up and test times for certain measurements, these cause amortized test cost to comprise a high percentage of the total cost of manufacturing semiconductor devices. In this work, we investigate a spatial correlation modeling approach using Gaussian process models to enable extrapolation of performances via sparse sampling of probe test data. The proposed method performs an order of magnitude better than existing spatial sampling methods, while requiring an order of magnitude less time to construct the prediction models. The proposed methodology is validated on manufacturing data using 57 probe test measurements across more than 3,000 wafers. By explicitly applying probe tests to only 1% of the die on each wafer, we are able to predict probe test outcomes for the remaining die within 2% of their true values.**

I. Introduction

As the final step in manufacturing RF devices, every fabricated circuit is tested against its design specifications before it is shipped to the customer. This testing is designed to capture devices that do not function properly due to manufacturing defects. Tests that screen for catastrophic faults, such as opens and shorts, are relatively inexpensive and straightforward to apply, but capturing parametric failures is considerably more difficult and typically requires RF specification testing.

While RF testing permits test engineers to capture the majority of devices failing due to parametric faults, RF automated test equipment (ATE) is very expensive, and test times can be quite long. Consequently, 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 [1], [2], or build machine learning models to learn classification boundaries separating passing and failing populations of devices [3], [4].

IEEE/ACM International Conference on Computer-Aided Design (ICCAD) 2012, November 5-8, 2012, San Jose, California, USA

In this work, we introduce a novel approach to reducing probe test cost via sparse spatial sampling and Gaussian process models [5] (also known as "kriging"). As we show in Figure 1, 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. As we demonstrate experimentally herein, via this test cost reduction approach, we are able to limit the total explicitly tested chips per wafer to 1%, while very accurately predicting the probe test outcomes for the remaining die. We observe prediction errors within an average/mean of 2% for 57 probe test measurements of more than 3,000 wafers.

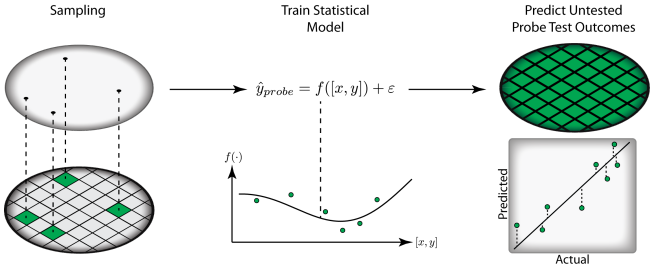

Fig. 1. Overview of Proposed Approach

The remainder of this paper is organized as follows. In Section II, we discuss existing work on statistical modeling for semiconductor manufacturing test cost reduction, and in Section III we discuss the spatial modeling approach taken in this work. Section IV introduces Gaussian process models and their relevance to semiconductor manufacturing. In Section V, we provide experimental results, and conclusions are drawn in Section VI.

II. Prior Work

Test cost reduction methodologies for RF device testing typically take a die-level view of the available semiconductor manufacturing data. By employing an exhaustively tested training set of devices in conjunction with a statistical model, such approaches realize test cost reduction by testing future

devices with alternative, lower-cost measurements (alternate test), or with only a subset of the existing test set (test compaction).

The alternate test approach [1], [2] employs non-linear multivariate adaptive regression spline (MARS) [6] models to correlate low-cost alternative test measurements against expensive probe RF tests during the final test process, thereby reducing overall test cost. This alternate test approach achieves dramatic test cost reduction, as the alternate test measurements are typically done at DC or low-frequency. However, this methodology comes with the cost of increased test escapes and yield loss; in the literature, reported errors are typically between 3% and 5%.

Test compaction-based approaches [3], [4] address the same problem, but by simply reducing the size of the original test set, favoring the lower-cost tests. Test compaction can be achieved either by retaining only non-RF tests [7] or by retaining a small number of RF tests [4]. In either case, the reduced test set is coupled with a statistical regression or classification model to identify failing device populations.

In both alternate test and test compaction, the costliness of semiconductor device testing is addressed by reliance on die-level statistical models to estimate the omitted test outcomes or to predict pass/fail labels. In contrast, in this work, we model probe variation spatially across *entire wafers*, thereby enabling us to leverage die-to-die correlations. This also codifies our *a priori* knowledge of semiconductor manufacturing processes, in that the majority of semiconductor fabrication steps introduce some type of cross-wafer variation.

For semiconductor manufacturing data, the most well-known spatial modeling methodology in the literature is an approach known as "Virtual Probe" [8], [9]. Virtual Probe models spatial variation via a discrete cosine transform (DCT) that performs a frequency domain projection from spatially sampled measurements. An important assumption of Virtual Probe is that spatial patterns of process variations are smooth, and consequently that they can be represented by a small number of dominant DCT coefficients at low frequencies [10]. The Gaussian process models that we employ in this work perform a more general projection via kernel functions. A complete empirical comparison of the proposed methodology and Virtual Probe is provided in the experimental results.

In [9], the authors extend Virtual Probe to address the same test cost reduction problem targeted by alternate test and test compaction. However, with Virtual Probe, the problem is addressed by performing wafer-level spatial sampling of expensive probe tests instead of completely removing them. This enables the authors to obtain a different tradeoff between test cost and the test escape or test error rate. In general, the spatial modeling problem addressed by Virtual Probe has similar properties to the problem addressed in this work. However, it takes a fundamentally different algorithmic approach, reasoning from the domain of compressed sensing rather than geostatistics.

A predecessor of this work can be found in [11], where the author lays the groundwork for applying Gaussian process models to spatial interpolation of semiconductor data based on Generalized Least Squares fitting and a structured correlation function. The computational method combines empirical data fitting and unconstrained optimization. In this work, we extend the key ideas of [11] by introducing modeling of radial variation and by providing an alternative derivation of Gaussian process model theory. We also demonstrate the first application of Gaussian process models to probe test spatial modeling, i.e. the problem addressed in [9] via the Virtual Probe methodology.

III. Proposed methodology

As noted in Section II, existing solutions to the test cost problem have typically eliminated all expensive probe tests and leveraged between-test correlations to estimate the unmeasured outcomes. In this work, we instead take a spatial modeling approach, leveraging correlations across spatial coordinates to provide probe test estimates. Instead of completely eliminating RF tests, we collect them for a sparse subset of die on each wafer. With the proposed methodology, we demonstrate that sparse spatial sampling can dramatically reduce test cost while incurring minimal additional test error.

To achieve this, we employ a spatial interpolation methodology targeted at providing accurate predictions for probe test outcomes at unobserved die locations. The spatial models generating such estimates are trained using a small subset of explicitly tested die, and consequently make a dramatic reduction in the cost of probe testing. Figure 1 depicts an overview of the proposed spatial interpolation methodology, and the following section provides a description of the statistical theory underlying the methodology, known as Gaussian process models.

IV. Gaussian Process Models

In this section, we articulate the theoretical underpinnings of Gaussian process models, as well as their usefulness for modeling spatial variation of wafer probe test parameters. Gaussian process modeling [5] is a spatial regression approach that models functions over Gaussian random fields based on sampled data. Such a regression approach is well-suited for spatial modeling of semiconductor test data, as it is extremely flexible in modeling data and imposes none of the traditional *a priori* assumptions about the underlying form of the generative function that tend to bias ordinary linear regression models.

Gaussian process models are birthed from the union of Bayesian statistics and the kernel theory of Support Vector Machines [12]–[14]. The fundamental concept underlying Gaussian processes is to model function outputs as drawn from a prior distribution with a fixed mean and a kernel-based covariance function.

We begin our discussion of Gaussian process model theory by considering the monolithic linear regression formulation $t = f(\mathbf{x}) + \varepsilon$, where $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{w}$ and ε represents independent and identically distributed (i.i.d.) additive noise.

With Gaussian processes, we do not presume the generative function $f(\mathbf{x})$ is of linear form in the original feature

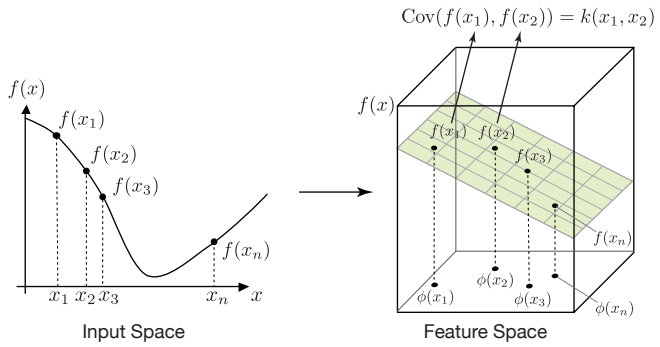

Fig. 2. Overview of Gaussian Process modeling

space, as shown by the one-dimensional input space curve on the left side of Figure 2. 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 n function evaulations $\{f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_n)\}$ over the coordinates is jointly Gaussian-distributed[2]. To derive a Gaussian process model for regression, we first consider a noise-free linear model, shown by the right side of Figure 2, which has the form:

$$t = f(\mathbf{x}) = \phi(\mathbf{x})^\top \mathbf{w} \tag{1}$$

where $\phi(\mathbf{x})$ is a function of the inputs mapping the input columns into some high dimensional feature space, shown by the bottom plane on the right side of Figure 2. For example, a scalar input $\mathbf{x}$ could be projected into the feature space: $\phi(\mathbf{x}) = (1, \mathbf{x}, \mathbf{x}^2)^\top$. We assign a Bayesian prior on the weights such that $\mathbf{w} \sim \mathcal{N}(0, \Sigma_p)$. As the realizations of the Gaussian process at points $\{f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_n)\}$ are jointly Gaussian, we can fully specify the Gaussian process with mean and covariance functions:

$$\mathbb{E}[f(\mathbf{x})] = \phi(\mathbf{x})^\top \mathbb{E}[\mathbf{w}] = 0, \tag{2}$$

$$\begin{aligned}\mathbb{E}[f(\mathbf{x})f(\mathbf{x}')] &= \phi(\mathbf{x})^\top \mathbb{E}[\mathbf{w}\mathbf{w}^\top]\phi(\mathbf{x}') \\ &= \phi(\mathbf{x})^\top \Sigma_p \phi(\mathbf{x}')\end{aligned} \tag{3}$$

A. Modeling Covariance with Kernel Functions

Consider the covariance function specified in Equation 3. Now, since covariance matrices are by definition positive semi-definite, we can redefine Σ_p as $(\Sigma_p^{1/2})^2$, and rewrite Equation 3 as:

$$\mathbb{E}[f(\mathbf{x})f(\mathbf{x}')] = \phi(\mathbf{x})^\top \Sigma_p \phi(\mathbf{x}') \tag{4}$$

$$= \phi(\mathbf{x})^\top (\Sigma_p^{1/2})^\top \Sigma_p^{1/2} \phi(\mathbf{x}') \tag{5}$$

We now introduce the parameter $\psi(\mathbf{x})$ by defining $\psi(\mathbf{x}) = \Sigma_p^{1/2}\phi(\mathbf{x})$, and subsequently rewrite the covariance of Equation 3 as:

$$\mathbb{E}[f(\mathbf{x})f(\mathbf{x}')] = \phi(\mathbf{x})^\top (\Sigma_p^{1/2})^\top \Sigma_p^{1/2} \phi(\mathbf{x}') \tag{6}$$

$$= \langle \psi(\mathbf{x}), \psi(\mathbf{x}') \rangle \tag{7}$$

[2]In this section, we adopt notation similar to [5] for convenience.

Crucially, this covariance function is formed as an inner product, permitting us to leverage the kernel trick [15] and express Equation 7 as a kernel function $k(\mathbf{x}, \mathbf{x}')$. In other words, the covariance between any outputs can be written as a function of the inputs using the kernel function without needing to explicitly computing $\phi(\mathbf{x})$. Many kernel functions exist, and any function $k(\cdot, \cdot)$ that satisfies Mercer's condition [12] is a valid kernel function. However, only a handful of kernels are commonly used. Among these common kernels, the most prevalent is the squared exponential, also known as the radial basis function kernel. In this work, we employed a squared exponential kernel of the form:

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2l^2}|\mathbf{x} - \mathbf{x}'|^2\right) \tag{8}$$

where l is some characteristic length-scale of the squared exponential kernel. Employing this kernel is equivalent to training a linear regression model with an infinite-dimensional feature space. Substituting our squared-exponential covariance function into the definition of the Gaussian process, we arrive at a Gaussian process formulation as:

$$t = f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}')) \tag{9}$$

The following section describes how to employ this process to derive predictive distributions, as well as how to manage the inclusion of additive noise in the model.

B. Training and Prediction

Suppose that we are provided a training set of n data points $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}^\top$ observed in an N-dimensional space, e.g., each vector in X is $\mathbf{x}_i = \{x_1, x_2, \ldots, x_N\}$, and a set of predictive targets, $\mathbf{t} = \{t_1, t_2, \ldots, t_n\}^\top$. Now, we wish to model the observed data as a noise-fee Gaussian process and define, as before, $y = f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$.

To derive the predictive distribution of this Gaussian process, we first write the joint distribution of the training set targets and a new test function value as:

$$\begin{bmatrix} \mathbf{t} \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K & \mathbf{k}_* \\ \mathbf{k}_*^\top & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix}\right) \tag{10}$$

Where $\mathbf{x}_*$ is a location we wish to extrapolate to, and where we have defined $K = K(X, X')$ as the matrix of the kernel function $k(\mathbf{x}, \mathbf{x}')$ evaluated at all pairs of training locations. We have also defined $\mathbf{k}_* = K(X, \mathbf{x}_*)$ as the column vector of kernel evaluations between the test point and the entire set training points, and lastly, $k(\mathbf{x}_*, \mathbf{x}_*)$ as the variance of the test function value at the observation point $\mathbf{x}_*$. With this distribution, we can condition the test function value on the observed data to obtain the predictive distribution (we omit the derivation for brevity):

$$\begin{aligned}f_*|X, \mathbf{t}, \mathbf{x}_* \sim \mathcal{N}(&\mathbf{k}_*^\top K^{-1}\mathbf{t}, \\ &k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^\top K^{-1}\mathbf{k}_*\end{aligned} \tag{11}$$

In this work, we primarily concern ourselves with point predictions, and so we use simply the distribution mean

$\bar{f}_* = \mathbf{k}_*^\top K^{-1}\mathbf{t}$ to generate a point prediction from the predictive distribution. This corresponds to decision-theoretic risk minimization [12], [13], [16] using a squared-loss function.

C. Regularization

To avoid overfitting, a technique known as regularization [16] is often employed in decision-theoretic empirical risk minimization. Gaussian process models handle regularization by considering the predictive tagets $\mathbf{t} = \{t'_1, t'_2, \ldots, t'_n\}$ as affected by additive noise such that $t'_i = t_i + \varepsilon$, where we make the usual i.i.d. assumptions about the additive noise $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$. To incorporate this into our Gaussian process model, we update Equation 9 to model additive noise in the observations:

$$y = f(\mathbf{x}) + \varepsilon \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}') + \sigma_n^2 \delta_{\mathbf{x},\mathbf{x}'}) \tag{12}$$

where $\delta_{\mathbf{x},\mathbf{x}'}$ is the Kronecker delta function. This, in turn, affects the joint distribution of Equation 10:

$$\begin{bmatrix} \mathbf{t} \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K + \sigma_n^2 I & \mathbf{k}_* \\ \mathbf{k}_*^\top & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix}\right) \tag{13}$$

as well as the predictive distribution:

$$f_* | X, \mathbf{t}, \mathbf{x}_* \sim \mathcal{N}(\mathbf{k}_*^\top (K + \sigma_n^2 I)^{-1}\mathbf{t}, \\ k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^\top (K + \sigma_n^2 I)^{-1}\mathbf{k}_* \tag{14}$$

resulting in a point prediction for new observations of $\bar{f}_* = \mathbf{k}_*^\top (K + \sigma_n^2 I)^{-1}\mathbf{t}$. This constrains the fitted model to avoid extreme predictions. For example, consider the univariate fit of Figure 3, shown with 4 monotonically increasing noise parameters $\sigma_n^2 = \{0, 0.0001, 0.01, 0.5\}$. The blue line is the fit model, the red dots are the original data, and the dotted red line is the true generative function. As this noise parameter increases, the model gradually flattens, and for very large σ_n^2, approaches a constant fit. Applying a model with a $\sigma_n^2 = 0$ is equivalent to the hypothesis that our observations are noise-free, this would be particularly useful to handle extreme excursions when modeling spatial correlation. As a practical matter, we have found empirically that $\sigma_n^2 = 0.1$ works well for our data. In the general case, this parameter should be adjusted to the particular application using a hold-out set of data.

In Figure 4, we show the effects of incorporating additive noise on example wafer data, with $\sigma_n^2 = \{0, 0.00001, 0.01, 0.1\}$. The bar on the right of each graph shows the measurement range on the wafer. As can be seen from the figure, modeling observations as noise-free leads to extreme variation in the model as it fits the response surface exactly through each point observation, and relaxing this constraint leads to smoother response surfaces.

D. Modeling Radial Variation

A key contribution of this work is the extension of Gaussian process modeling over Cartesian coordinates to a joint Cartesian-radius space, capturing our intuition that wafer variance is often radial. By including a radius feature, we

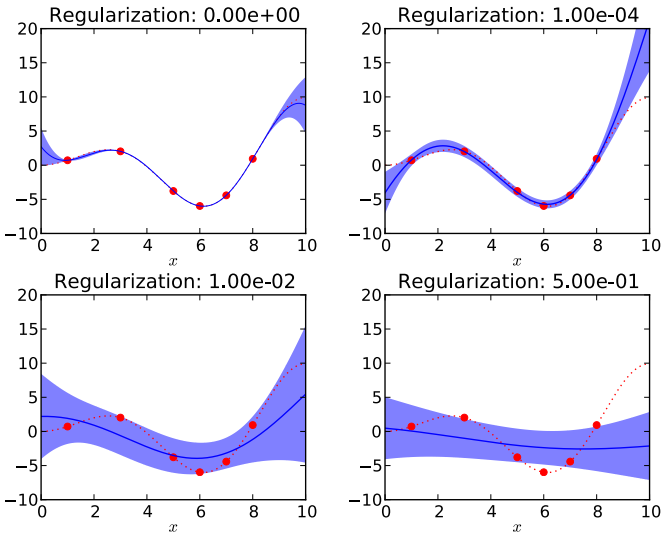

Fig. 3. Regularization Example

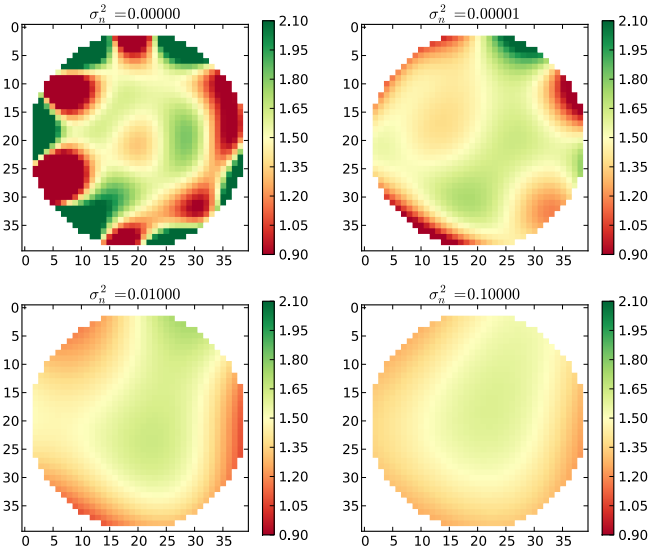

Fig. 4. Wafer Regularization

canonicalize the notion that any set of die drawn from a wafer-centered ring should present similar probe test measurement profiles.

An advantage of using Gaussian process regression is the ability to apply a Gaussian process over arbitrary index sets. Thus far, we have been describing a Gaussian process implementation that estimates probe test parameters over a 2D Cartesian plane, but we are free to use any other field. As noted above, many parameters will manifest radial variation patterns due to the physical realities of semiconductor manufacturing. To accommodate this in our Gaussian process model, we can simply update our coordinates from $\mathbf{x} = [x, y]$ to include a radius $r = \sqrt{x^2 + y^2}$:

$$\mathbf{x} = \left[x, y, \sqrt{x^2 + y^2}\right]$$

Now, applying the Gaussian process regression model over this space will result in a model that takes radial variation patterns into account. In Figure 5, we show the impact on the prediction outcomes.

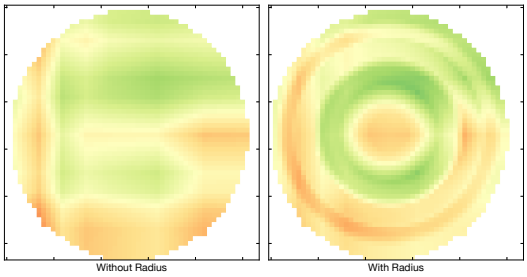

Fig. 5. Radial Modeling

E. Gaussian Process Models for Probe Test Interpolation

To spatially interpolate probe test measurements across each wafer, 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 $[x, y]$ die locations. By approaching spatial modeling on a per-wafer basis, we sidestep the need for the "median polishing" methodology of [11].

We capture the effectiveness of our proposed methodology by recording the percentage prediction error of our statistical model on each measurement and each wafer. Consequently, the output of our Gaussian process implementation is a $j \times k$ matrix E composed of prediction errors ϵ_{jk}, where ϵ_{jk} is the prediction error of the model for the j-th wafer and the k-th probe test parameter.

To train the Gaussian process model for a particular wafer j, we collect probe test data from a 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. Specifically, consider observations of the k-th probe test parameter $t_k^{(i)}$ measured at training sites $i \in \{1, 2, \ldots, n_{train}\}$, where each site has an associated $\mathbf{x}_i = [x, y]$ location consisting of the Cartesian coordinates of the die. After training the model on this data, it is then used to predict probe test outcomes for the remaining die and produce an estimate:

$$\hat{t}_k^{(l)} = \bar{f}_* = \mathbf{k}_*{}^\top (K + \sigma_n^2 I)^{-1} \mathbf{t} \tag{15}$$

This process is then repeated for every die in the test set. We compare these predictions to the true probe test outcomes for the test set to compute test error, and record the mean absolute percent error across all predictions:

$$\epsilon_{jk} = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \left| (\hat{t}_k^{(i)} - t_k^{(i)}) / t_k^{(i)} \right| \tag{16}$$

Thus, ϵ_{jk} represents the mean percent error of predicting the k-th probe test parameter for all test set die locations on a particular wafer j. Applying the model in this fashion for all

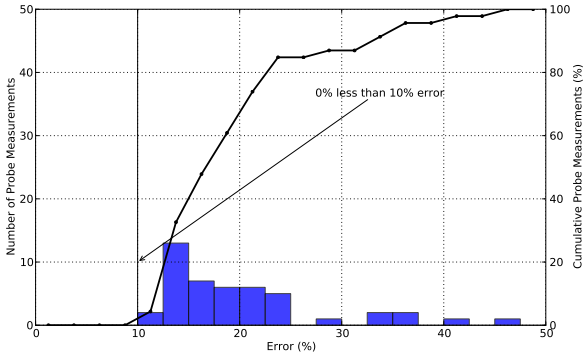

Fig. 6. Virtual Probe prediction error across all wafers

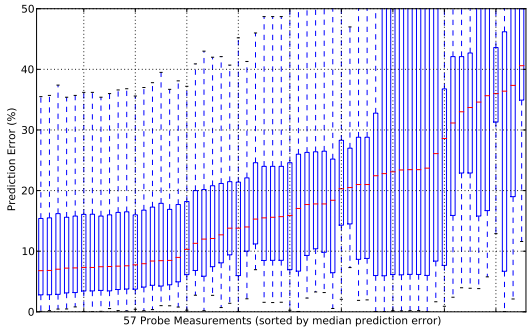

Fig. 7. Virtual Probe prediction error for each probe test measurement

wafers and all probe test parameters, we populate the matrix E that completely characterizes the performance of the spatial interpolation models on the dataset at hand. We can also summarize mean prediction error for a particular probe test parameter by computing the mean error over all wafers:

$$\epsilon_k = \frac{1}{N_{wafers}} \sum_{i=1}^{N_{wafers}} \epsilon_{ik} \tag{17}$$

V. Experimental Results

In this work, we demonstrate results on probe test data from high-volume semiconductor manufacturing. The device under consideration is an RF transceiver with multiple radios built in a 65nm technology. 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 randomly sampled on each wafer were used 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, and the mean prediction errors were computed through Equation 16, as described in Section IV-E. Consequently, the resultant matrix of prediction errors E was of dimension $3,499 \times 57$.

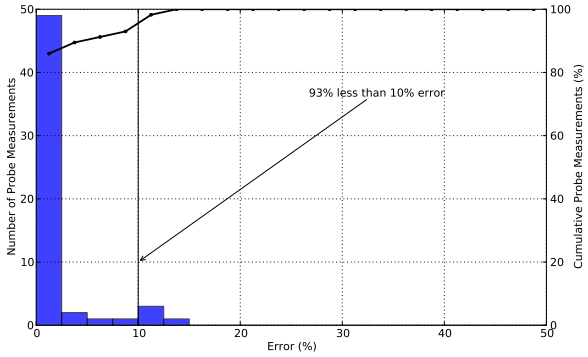

Fig. 8. Gaussian process model prediction error across all wafers

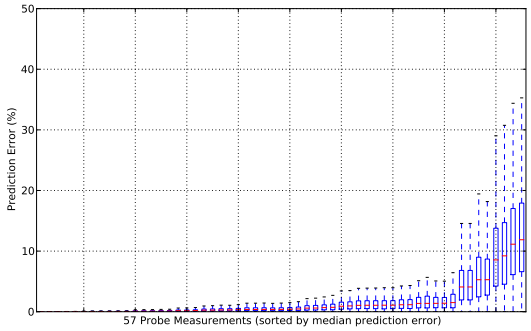

Fig. 9. Gaussian process prediction error for each probe test measurement

A. Virtual Probe Results

To provide a reference for the performance of the proposed methodology, we first applied the Virtual Probe approach of [9] to the dataset at hand. In Figure 6, we present a histogram of the experimental prediction errors for Virtual Probe. The superimposed black line on the histogram represents the cumulative number of probe test measurements. Virtual Probe performs moderately well, with global mean prediction error of 21.0%. However, several of the probe test measurements observe high prediction error, exceeding 40%.

In Figure 7, we present an overview of the Virtual Probe prediction errors with 10%–90% error bars shown for all 57 probe test measurements, sorted by median Virtual Probe prediction error. The x-axis corresponds to probe test measurements, and the y-axis shows the prediction error, in percent, incurred by Virtual Probe. Again, note that for many of the probe test measurements, the prediction error is quite high.

B. Gaussian Process Modeling Results

The proposed Gaussian process modeling approach was evaluated via the same metrics as Virtual Probe. In Figure 8, a histogram of the mean probe test prediction errors is presented, again as computed via Equation 17. Note that, as in Figure 6, the superimposed black line represents cumulative probe tests. Evidently, the Gaussian process model prediction errors are substantially lower than those observed by Virtual Probe, with more than 93% of the probe test measurement predictions exhibiting less than 10% error.

In Figure 9, we present the Gaussian process model prediction errors with 10%–90% error bars. By comparing the observed errors to those of Virtual Probe presented in Figure 7, we observe that the proposed methodology generates substantially lower prediction errors in the test set, and with much tighter error bars. Note that the widths of the error bars are quite small, indicating that the prediction 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.

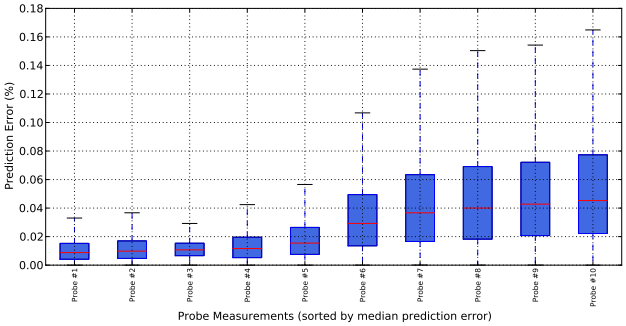

Fig. 10. Prediction error for best 10 probe test measurements

Figure 10 presents a zoomed-in version of Figure 9 focusing on the best-predicted 10 probe measurements, e.g., the left-hand side of Figure 9. The x-axis shows these 10 probe test parameters, which include a mixture of RF performances; the y-axis presents percent error. For these 10 measurements, the mean prediction error is less than 0.05%, indicating that we can very successfully model these parameters.

In Figure 11, we present another zoomed-in version of Figure 9, in this case focusing on the worst-predicted 10 probe measurements, or the right-hand side of Figure 9. Even in the worst case, our prediction errors are typically below 10%.

C. Comparison to Virtual Probe

In Figure 12, we present a comparison of the two methodologies, with Virtual Probe set as the baseline at 0%. 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 I. The per-wafer training and prediction time for Virtual Probe and Gaussian process models is also presented in Table I; the proposed methodology is extremely fast and requires less than a second to complete the full train-predict cycle for

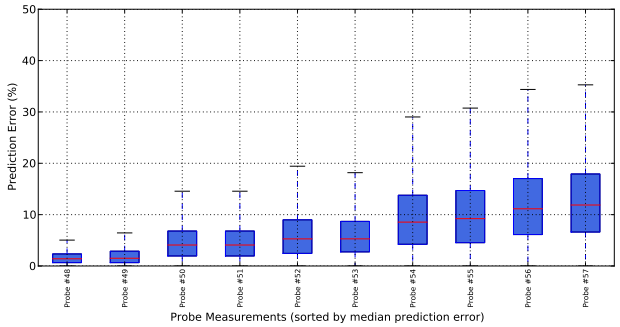

Fig. 11. Prediction error for worst 10 probe test measurements

an entire wafer. Notice that the measurements used to train the models are specification probe measurements required to release the products, thus no additional costs are incurred to train the model. The timing measurements were collected on a 2010 Core i5 2.4GHz MacBook Pro, and 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, which can significantly reduce the time and cost of applying probe specification measurements.

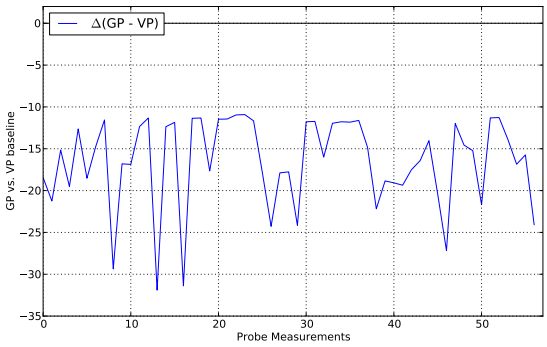

Fig. 12. Relative Error of Gaussian process (GP) models vs. Virtual Probe (VP)

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

TABLE I
COMPARISON OF VIRTUAL PROBE & GAUSSIAN PROCESS MODELS

VI. CONCLUSION

We have demonstrated a spatial correlation modeling approach which dramatically reduces probe test cost for RF devices. By sparsely sampling probe tests and extrapolating to untested die locations, our proposed methodology avoids the requirement for dense application of costly probe tests. As demonstrated on more than 3,000 high-volume manufacturing 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 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.

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 first author is supported by an IBM/GRC (Global Research Collaboration) graduate fellowship.

REFERENCES

[1] 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.

[2] 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.

[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 VLSI Test Symposium*, 2007, pp. 9–14.

[5] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.

[6] J. H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 1991.

[7] N. Kupp, P. Drineas, M. Slamani, and Y. Makris, "Confidence estimation in non-RF to RF correlation-based specification test compaction," in *Proc. of the European Test Symposium*, 2008, pp. 35–40.

[8] X. Li, R. R. Rutenbar, and R. D. Blanton, "Virtual probe: A statistically optimal framework for minimum-cost silicon characterization of nanoscale integrated circuits," in *Proc. of International Conference on Computer-Aided Design*, 2009.

[9] H.-M. Chang, K.-T. Cheng, W. Zhang, X. Li, and K. Butler, "Test cost reduction through performance prediction using virtual probe," in *Proc. of International Test Conference*, 2011.

[10] W. Zhang, Xin. Li, F. Liu, E. Acar, R. A. Rutenbar, and R. D. Blanton, "Virtual probe: A statistical framework for low-cost silicon characterization of nanoscale integrated circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, no. 12, pp. 1814–1827, 2011.

[11] F. Liu, "A general framework for spatial correlation modeling in VLSI design," in *Proc. of Design Automation Conference*, 2007, pp. 817–822.

[12] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.

[13] V. Vapnik, *Statistical Learning Theory*, John Wiley and Sons, Inc., 1998.

[14] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, 2001.

[15] M. A. Aizerman, E. A. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning," in *Automation and Remote Control*, 1964, vol. 25, pp. 821–837.

[16] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2001.