Yield Prediction Under Non-Standard Data Distributions

Suraj Rao, Sharad Saxena, Pushkar Apte, P. K. Mozumder,
Joseph Davis, Richard Burch, and Karthik Vasanth
Texas Instruments Incorporated, P.O. Box 650311, MS 3704, Dallas, TX 75265

I. INTRODUCTION

Manufacturability is an important issue for deep-submicron micro-electronic devices. Yield, defined as the fraction of the total manufactured devices that meet all performance specifications, is an important component of manufacturability. Both yield evaluation and design for yield improvement require accurate characterization of performance variation through its distribution. For parametric yield, performance variation results from process variations during manufacturing, such as variations during gate-patterning or gate-oxidation in MOS devices. For catastrophic yield, performance variation results from the variation in the number, size, and shape of the particles on a wafer. In either case, an accurate estimation of the distributions of the performance variables is essential for yield estimation and optimization.

A typical approach for estimating a performance distribution is to assume an appropriate form for the parametric distribution and estimate the parameters from historical data or designed experiments. For example, particle numbers are assumed to follow a Poisson distribution, while most continuous performance variables are assumed to follow a normal distribution.

The assumption of a fixed parametric form, known a-priori, is often not valid. This is especially true with the trend towards increasing wafer sizes where variation contribution from non-random spatial patterns cause significant departure from the normal distribution assumption [2]. In this situation, performance distributions can take arbitrary or non-standard shapes and are not easily parameterizable; these non-parametric distributions make rapid yield calculation a difficult task.

In this paper, we propose a method that allows yield prediction under arbitrary shaped distributions. Furthermore, we provide a compact representation of these non-parametric distributions to allow rapid yield calculation for manufacturability methods that attempt to design in robustness to manufacturing variations [3]. Section II sets-up the yield prediction problem and discusses the traditional approaches and the required generalization to calculate yield under non-standard data distributions. Section III describes a design of experiment for the salicide process where we first encountered the problem of non-parametric distributions. Section IV discusses a method to extract and represent the non-parametric reference distribution from experimental data. Section V presents an
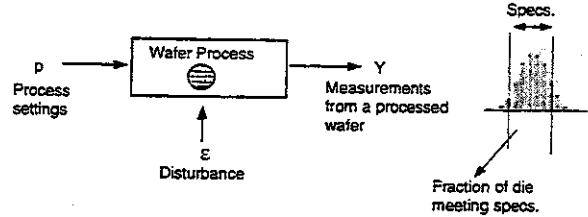

Fig. 1. An abstract manufacturing process; process yield is a function of the distribution of output Y.

accuracy comparison for probability calculation based on normal distribution assumption with that based on a non-parametric approach. Section VI summarizes the contribution of this paper.

II. PROBLEM FORMULATION

Figure 1 represents an abstraction of a integrated circuit manufacturing process. In this process a silicon wafer is subjected to process setting $\mathbf{p}$. The process is also subject to inherent manufacturing disturbances denoted by ϵ. The output from the process is the performance measurements Y made on a randomly selected die on the wafer. Process yield, or the probability of meeting target specifications given $\mathbf{p}$ is denoted by

$$P(Y \in T | \mathbf{p}) \qquad (1)$$

where T is the target space or specification limits for the performance measurement. This probability is calculated with respect to the probability measure of the output distribution and is a function of the assumptions made on the shape of the distribution. There are at least two standard approaches to characterize the mean, the variance, and the form of the output distribution.

In the first approach, the output Y is modeled as a function of process settings plus an independent random disturbance, that is,

$$Y = f(\mathbf{p}) + \epsilon. \qquad (2)$$

The function f is typically approximated by a second order response surface model (RSM) built from data from a designed experiment. The disturbance ϵ is assumed to be independent and identically normally distributed with zero mean and a constant variance σ^2. The constant variance σ^2 is estimated from experimental data. Under this formulation, the distribution

of performance Y at any process setting $\mathbf{p}$ is given by

$$Y \sim N(f(\mathbf{p}), \sigma^2)$$

and the probability calculation in (1) becomes a simple matter.

In a second approach, one retains the independence and normal distribution assumption on ϵ but assumes that the variance itself is also a function of process settings $\mathbf{p}$. Specifically, the variance of the within-wafer measurements is modeled as a function of the process settings, that is

$$\log(\sigma^2) = g(\mathbf{p}) + \zeta$$

where g is also approximated by an RSM. ζ is the additive error assumed to be distributed normally with mean zero and a constant variance denoted by α^2 [4]; this variance is also estimated from experimental data. The log transformation ensures that the non-negative variance-response is mapped on to the real line. Under these assumptions, it can be shown that

$$\widehat{\sigma^2}(\mathbf{p}) = \exp(g(\mathbf{p}) + \frac{\alpha^2}{2})$$

is an unbiased estimator of σ^2 for each process condition $\mathbf{p}$. Now the probability calculation in (1) can be carried out easily by assuming

$$Y \sim N(f(\mathbf{p}), \widehat{\sigma^2}(\mathbf{p})).$$

Both of the above approaches assume that the error distribution is independent and identically normally distributed. These approaches may be generalized to other parametric distribution forms and still permit yield calculation. But what if the output or performance distribution does not fall in the parametric class?

Figure 2 shows the spatial pattern across a wafer for one of the electrical parameters denoted by Y, and is based on measurements made on 24 die-locations on the wafer. The non-random pattern is typical of many electrical measurements. The figure also shows the frequency histogram of the measurements. For the purpose of yield calculation, we are interested in predicting the fraction of die-measurements that meet specifications. We make two important observations about the histogram. First, for yield calculation, the histogram captures the total effect of both the deterministic pattern across the wafer and the random variations in the measurements. Second, though the individual measurements on the die are not independent of each other, the histogram captures the total variation that will be seen across the wafer. Conceptually, this is equivalent to separating all the die on the wafer, shuffling them in an urn, and taking a measurement from a randomly selected die. As far as yield calculation is concerned, this is the only distribution
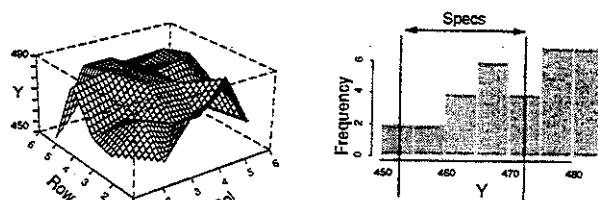

Fig. 2. A wafer-map showing within-wafer variation of electrical data and the corresponding frequency histogram for the data.

of interest; individual die-locations are of no relevance for the final yield calculation.

The above histogram or distributions can have arbitrary shapes; the shape will depend on both the deterministic and the random component of within-wafer variation. How do we perform yield calculation based on these arbitrary shaped distributions? In the next two sections, we address various aspects of this question. Specifically, in Section III we further explore the possibility of representing the within-wafer distribution by a parametric distribution. In Section IV, we propose a method to improve our confidence about the shape of the within-wafer distribution instead of relying on a distribution based on 24 die measurements as suggested in Figure 2. This section also discusses a compact representation of this distribution for use in a TCAD framework.

III. A PROCESS EXAMPLE WITH NON-PARAMETRIC DATA DISTRIBUTION

One of the steps in MOS device processing is the formation of titanium silicide to make contacts to silicon or polysilicon in an integrated circuit. Titanium silicide reduces the contact resistance when metal contacts are made later to these regions. The process involves the implantation of ions to form a diffusion region followed by a deposition of a layer of titanium to form titanium silicide. The critical performance of interest is the contact resistance between the silicide and the diffusion region.

An experiment was designed to study the silicide contact resistance quality as a second order quadratic function of 4 input variables: the source/drain implant dose, the titanium thickness, the silicide anneal time, and the silicide anneal temperature. These inputs were selected based on prior engineering knowledge and previous screening experiments. The design of experiment for the 4 input variables consisted of a G-optimal design with 20 independent runs and 4 replicate runs. A batch of 24 wafers was processed through a standard CMOS process flow with the silicide settings as suggested by the above design.

After processing the 24 wafers, contact resistance measurements were available from 24 die locations from each of the 24 wafers. From the within-wafer

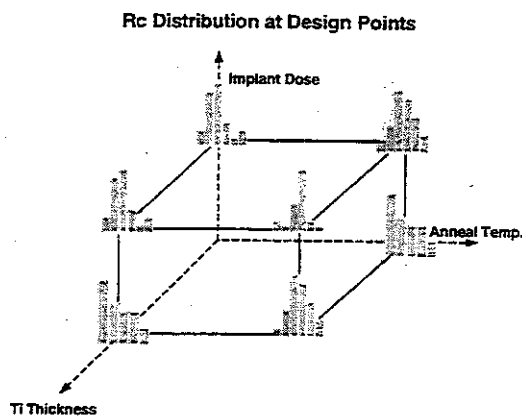**Rc Distribution at Design Points**

Fig. 3. Distribution of contact resistance as a function of process settings.

contact resistance data, it was clear that both the mean and variance of the within-wafer measurements were a function of the process settings. Figure 3 shows the shape or form of the within-wafer distribution of contact resistance measurements for 8 of the sample design points that received identical silicide anneal time. The actual values of the contact resistance are not relevant and are not shown to preserve proprietary information.

Based on the the data from the experiment, it is possible to build RSMs for both the mean and within-wafer variance of the contact resistance as a function of the process settings. But to calculate the probability given in (1) at any interior point in the process space, we need to characterize the underlying shape of the frequency distribution from the available data. The idea is to test if the within-wafer data come from one of the well known statistical distributions. We performed this test using the Kolmogrov-Smirnov (KS) statistic [1].

Briefly, the KS test allows us to determine if two random variables have similar probability distributions and is based on the comparison of their cumulative distribution functions (CDF). The CDF of a random variable X is defined as $F(x) = P(X \leq x)$ where P is the probability density function of X. If the actual underlying distribution is unknown, and instead, only a sample of observations are available from a distribution, the CDF is estimated as being equal to the proportion of observations in the sample that are less than or equal to x. Now let X_1 and X_2 be two random variables with CDFs F_1 and F_2. We wish to test whether the two random variables have similar probability distributions, that is, whether F_1 and F_2 are statistically equivalent. This equivalence can be tested using the KS statistic defined as

$$D = \sup_x |F_1(x) - F_2(x)|.$$

If the test statistic is greater than a critical value then the assumption of equivalence is rejected. The critical

values are read-off from the tables of quantiles of KS test-statistic distribution [1].

In our problem, we wish to determine whether the within-wafer measurements have a distribution similar to that of, say, a normal distribution. In this case, we let F_1 denote the empirical CDF constructed from the within-wafer measurements from one of the wafers and F_2 the CDF of a normal distribution with the mean and variance as those estimated from the within-wafer measurements from that wafer. We performed the KS test for each of the 24 wafer separately and found that 22 of these wafers showed evidence to reject the normality assumption at 0.1 significance level. As a result, the normal distribution assumption was found to be untenable.

Furthermore, we performed a series of Box-Cox transformations to improve the normality of the data but were unable find a single common transformation that simultaneously improves the normality of the within-wafer measurements for each of the settings. We also performed the above tests with two other candidate parametric non-normal distributions, namely, the t and the Chi-square. Neither of the two were able to fit the distribution of the within-wafer measurements simultaneously for all process settings.

Although only a small set of transformations and parametric family of distributions were attempted to capture the underlying distribution shape, we found the procedure to be time consuming and laborious. Even if one is willing to expand the candidate sets to other transformations and parametric distributions, there is no guarantee that the procedure will find a closed form distribution to capture the within-wafer measurements. Is there a better method to represent the underlying within-wafer variation? In the next section, we take a non-parametric approach to characterize the within-wafer distribution.

IV. CAPTURING VARIATION THROUGH NON-PARAMETRIC DISTRIBUTIONS

In this section, we propose a non-parametric distribution to capture the underlying shape of the distribution of the within-wafer measurements at each the design points. We first describe the method of construction of the proposed distribution and then use a cross-validation approach to check the validity and adequacy of the proposed distribution to represent the distribution of the within-wafer measurements.

We make one assumption about the underlying distribution of the within-wafer measurements. We assume that whatever be the *form* of the underlying distribution, it does not vary with changes in the process settings and that only the mean and the variance of this distribution are functions of the process conditions. This assumption is similar to the one made in the parametric modeling framework. For example, if the form of the distribution is assumed to be normal,
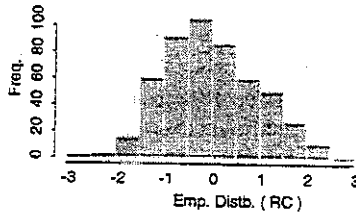
Fig. 4. The reference distribution captures the underlying shape of the within-wafer variation.

all the within-wafer measurements are assumed to remain normally distributed at all design points; only the distribution mean and variance will be functions of process settings.

Under the above assumption, there exist a common underlying form for the distribution of within-wafer measurements. This assumption also implies that each set of within-wafer measurements are samples from this underlying distribution with only the mean and variance being the function of process settings. We obtain the form of the underlying distribution by accumulating the shape evidence from the within-wafer measurements from each of the wafer. Specifically, the within-wafer measurements are collected from each wafer and the measurements are centered and scaled with the wafer mean and standard deviation; this ensures that the standardized data have zero mean and unit standard deviation. Next, the standardized measurements from the individual wafers are aggregated together to build a common distribution or a histogram. This histogram or *reference distribution* is shown in Figure 4. This reference distribution is thought to capture the underlying form of the distribution of the within-wafer measurements. We performed a KS test to confirm that the reference distribution is significantly different from a normal distribution at 0.01 significance level. This reference distribution is thought to capture the underlying form of the distribution of the within-wafer measurements. The standardized reference distribution is then represented by its empirical CDF and all probability calculation can be made from this distribution.

But how do we test the initial assumption of a common underlying distribution for the within-wafer measurements, and the claim that the underlying distribution is the reference distribution? We test the assumption and the claim by the method of cross-validation.

Consider the within-wafer measurements from wafer $i, i = 1, \ldots, 24$. Let F_i represent the empirical CDF obtained from these measurements after centering and scaling them with the mean and standard deviation from the ith wafer. Let F_{-i} denote the empirical CDF of the reference distribution without the measurement contribution from the ith wafer. We de-

fine the KS statistic

$$D_i = sup_x |F_i(x) - F_{-i}(x)|$$

for $i = 1, \ldots, 24$. For each D_i, we perform a statistical significance test at 0.01 level. We found that except for 2 of the wafers, the distribution of the individual within-wafer measurements was statistically similar to the the reference distribution. Thus if we are willing to ignore the anomaly from the two wafers, the reference distribution is a good approximation of the form of the distribution; furthermore, the form is independent of the processing conditions. Note that in the cross-validation approach of testing for similarity of distribution of the within-wafer measurements with that of the reference distribution, we first removed the effect of the ith wafer from the reference distribution so as not to bias our test.

Though the reference distribution does not have a closed parametric form, we can estimate the probabilities in (1) using the empirical CDF for this reference distribution. The calculation is performed in two steps. First, the RSM models for the mean and variance are used to calculate the mean and standard deviation of the response as a function of process settings. Now the measurements in the reference distribution are centered and scaled with these values. The new data are used to build the empirical distribution function of the response at the given process settings. The required probabilities can be easily calculated from the empirical distribution function.

In the next section, we illustrate the method of probability calculation and compare the accuracy of our approach to that of the calculation based on the assumption of a normal distribution.

V. COMPARISON OF METHODS

Suppose we are interested in finding the probability that R_c, in the scaled units, is less than 2.75 $\Omega - \mu m$ for a given process condition. We can perform this calculation for any process condition in the design space. For validation purpose, we choose a process condition that corresponds to one of the design points from our experiment for the salicide process. Figure 5 shows the scaled within-wafer distribution. From the R_c measurements, the fraction that fall below 2.75 $\Omega - \mu m$ is $11/23 = 0.4783$. Using this as the benchmark value, we can compare the accuracy of the traditional normal approximation to that based on our reference distribution approach. The comparison is summarized in Table 1. The mean and standard deviation of the measurements from this wafer is 3.08 and 1.33 $\Omega - \mu m$ respectively. If one makes the normal assumption, then $P(0 < R_c < 2.75 | normal) = 0.3992$ giving an error of about 17%. On the other hand, when we use the same mean and standard deviations to center and scale the reference distribution, the fraction of measurements lying in the same interval is 0.4596 giving an error of 4%.
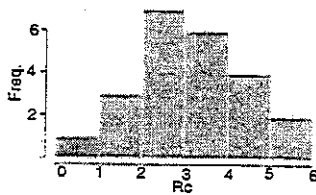
Fig. 5. Distribution of within-wafer scaled R_c measurements for a sample wafer.

TABLE I
ACCURACY COMPARISON

Method	$P(0 < R_c < 2.5)$	Percent Error
Observed Value	11/23= 0.4783	0.0
Normal Approx.	0.3992	17.0
Empirical CDF	0.4596	4.0

The above example illustrates the method of using the empirical CDF for probability calculation. In this example, the probability calculation uses only a small portion of the reference distribution and the corresponding normal distribution to illustrate the relative accuracy. This example is is not used to claim the appropriateness of the reference distribution approach. The appropriateness was determined using the more powerful KS test which uses all parts of the reference distribution during comparison with a candidate distribution.

VI. DISCUSSION

The reference distribution may be used for design for manufacturability. For instance, one may be interested in finding a process condition that will maximize the fraction of die that would give R_c measurements below a certain value. Since our method allows us to predict the distribution at each point in the design space, the problem reduces to an optimization problem where each function evaluation is a probability calculation based on the reference distribution.

There are several advantages to the empirical CDF approach. First, one can represent arbitrary shaped distributions which may even incorporate within-wafer spatial dependency information. In our example, we used parametric or continuous data for yield calculation. Instead, we can replace the data with count or catastrophic yield data. Since the approach is independent of the shape of the distribution, the same techniques may be used for yield calculation. Second, the class of parametric distributions assumed in the traditional method is a special case of our general approach; we can still represent a parametric distribution by its empirical CDF. Third, the complete predictive distribution at any interior point in the de-

sign space can be characterized compactly by three quantities, an empirical CDF function, a mean estimate and a variance estimate. As a result, the compact representation can become part of TCAD tools used for design for manufacturability.

REFERENCES

[1] Deshpande, J.V., Gore, A. P., and Shanubhogue, A., " Statistical Analysis of Non-normal data", *Wiley Eastern Limited*, New Delhi, India.

[2] Kibarian, J. K., and Strojwas, A. J., (1991) "Using Spatial Information to Analyze Correlations Between Test Structure Data", *IEEE Transaction on Semiconductor Manufacturing*, Vol. 8, No. 2.

[3] Saxena, S., et el. "Method for the Design of Microelectronic Devices and Process Flows for Manufacturability", *Proceedings of SPIE; Microelectronic Device Technology*, Vol. 3212, pp 18, October 1997.

[4] Vining, G.G., and Schaub, D., " Experimental Design for Estimating both Mean and Variance Functions", *Journal of Quality Technology*, Vol. 28, No. 2, April 1996.