Model-Free Estimation of Some Yield Metrics in Integrated Circuit Fabrication

David J. Friedman
Engineering Research Center
AT&T Bell Laboratories
Princeton, New Jersey

Mark H. Hansen
Statistics Research
AT&T Bell Laboratories
Murray Hill, New Jersey

Vijayan N. Nair
Department of Statistics
University of Michigan
Ann Arbor, Michigan

David James
Statistics Research
AT&T Bell Laboratories
Murray Hill, New Jersey

Abstract

This paper describes a model-free method of estimation for two important metrics used to track the integrated circuit fabrication process: D_0, the random defect density, and Y_0, a measure of (the lack of) defect clustering. Unlike previous methods discussed in the literature, our approach makes extensive use of the spatial information in wafer map data to estimate these metrics at the wafer level. Simulation studies show that our procedure is superior to the time-honored windowing technique. The potential applications of this methodology for quality improvement activities are outlined.

1. Introduction

Over the past 30 years, various metrics have been proposed to track the fabrication process of integrated circuits (IC's). In this paper, we consider techniques for estimating two such quantities: D_0, the number of defects per square centimeter; and Y_0, a measure of (the lack of) defect clustering. Metrics such as these have been used by product and process engineers for two principal purposes: i) to better calculate capacity and cost estimates by constructing models to predict future yields; and ii) as a tool for comparing yields of products of differing size and complexity. Our interest in D_0 and Y_0, on the other hand, is from a quality improvement point of view. It has been our experience that routine identification and analysis of clustered defects from wafer map data can lead to the exposure of root cause problems, and hence, represents a cost-effective approach to improving yield (Friedman et al [1995]).

To make this more precise, assume that the pattern of defective chips on a given wafer is in fact a superposition of two mechanisms. Acting in isolation, the first generates defective chips uniformly across the wafer, while the second produces clusters or large-area defects. In Figure 1, we illustrate how these two mechanisms combine to yield an observed wafer map. In each of the three maps in this figure, defective chips are colored black and fault-free chips are colored white. Now, given a wafer map, we would like to estimate the size of the large-area defects, if they exist, and the defect density associated with the random defects;

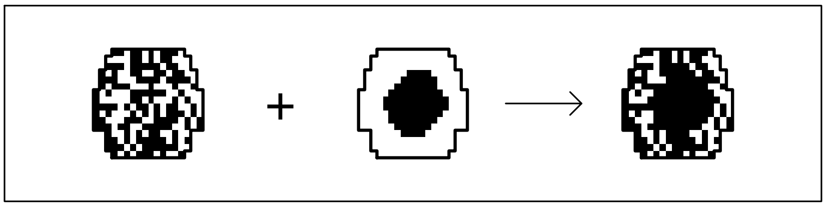

Figure 1. *An idealized view of two destructive forces working independently to create the observed wafer map.*

in essence, reversing the arrow in Figure 1. In this paper, we describe a new technique for accomplishing this task by first estimating which defective chips, if any, can be assigned to a large-area defect. By actually isolating clusters, this procedure can be used in conjunction with yield improvement efforts by providing a graphical summary of the patterns of defective chips present within a lot of wafers (Friedman et al [1995]). For the purpose of this paper, however, we will focus on how well our technique performs at judging the *extent* of spatial clustering; that is, how well we can estimate Y_0. As a benchmark, we will compare our technique with the windowing procedure discussed in the IC fabrication literature. The applications of our methodology to process improvement efforts will be briefly discussed in the last section of the paper.

1.1 Why do clustered defects occur?

Clusters like the one exhibited by the middle wafer in Figure 1 are typically classified as either particle- or process-related defects. Particulate clusters are often caused by machine failures of some kind. Excess vibration, for example, might liberate enough particles to cause all the chips in some contiguous region to fail. In general, these particle-related clusters tend to be assignable to individual machines or processes. Similarly, process-related clusters are often caused by a process step not meeting its required specifications. For example, an uneven growth of oxide on a wafer could cause all the chips in a certain sector of the wafer to fail. In either case, by routinely monitoring wafer map data for spatial clustering, we can identify potential problems in the fabrication process and thereby improve yield.

On the other hand, spatially random defects like those exhibited by the first wafer in Figure 1 are most often the product of the clean room environment. Random defect densities tend to rise and fall with the overall cleanliness of the clean room, and can only be reduced through a program of long term, gradual improvement or expensive equipment overhaul. We do not mean to imply, however that the intensity of the spatially random component of yield is devoid of information for yield improvement. On the contrary, process-related problems that result in excess wafer roughness, for example, can manifest themselves in low yielding wafers that do not exhibit any signs of spatial clustering. Therefore, we believe that both components of wafer yield should be routinely monitored for irregularities.

1.2 Relationship to Traditional Yield Models

Before defining the parameters D_0 and Y_0 precisely in our context, we briefly review the literature on yield models where these concepts have been discussed extensively. The early yield models were based on the Poisson distribution: if D_0 denotes the average number of defects per unit area and A denotes the (active) area of the chip, then the probability that the chip has k defects under the Poisson model is

$$P(k) \;=\; (D_0A)^k \exp(-D_0A)/k! \tag{1}$$

In particular, the expected yield (the probability that a chip is fault-free) is given by

$$Y \;=\; \exp(-D_0A)\,. \tag{2}$$

Unfortunately, as chip sizes increased the Poisson model proved to be inadequate, underestimating the yield for products with large chip areas (Cunningham [1990]). These poor predictions stemmed from the fact that the Poisson model simply did not fit the observed yield data. For example, while the mean and variance of random variable having a Poisson distribution are equal, the variance of observed yield data was often found to be much larger than the mean. This led to the development of a variety of new models, most of which fall into two categories aimed at addressing two different kinds of defect clustering.

One type of model attempts to segment defects into two classes: i) large-area defects, large relative to the area of the chips, and ii) small area defects which are random events occurring uniformly on the wafer. Under this framework, if Y_0 denotes the proportion of chips not affected by large-area defects, the expected yield for a randomly chosen chip is

$$Y \;=\; Y_0 \exp(\,-D_0A\,)\,. \tag{3}$$

While this is similar in spirit to the set-up in Figure 1, the definition of what constitutes a large-area defect varies considerably in the literature. Following Murphy [1964], some engineers view a large-area defect to be one which is independent of chip size and one in which every chip within the area is defective. They then interpret D_0 as a surrogate for particle based problems and Y_0 as a proxy for process failures. Unfortunately, using this classification, it is virtually impossible to distinguish between a Y_0 failure and a D_0 failure solely on the basis of wafer map data. Many fab engineers believe that to truly understand the nature of an observed fault, some micro-inspection is required, making the strict particle- versus process-related failure separation an extremely time consuming proposition. We believe, however, that this distinction may not be relevant from a yield improvement standpoint. Consistent particulate failures that occur with sufficient intensity to produce large clusters of defective chips can be important indicators of process problems.

The second class of models attempts to account for the fact that the density of the spatially random defects varies both within a wafer and from wafer to wafer. To capture this variation, a distribution for the defect density, $f(D)$ with average density D_0, was introduced. This leads to a family of compound Poisson models for the defect distribution. In particular, the overall yield is given by

$$Y = \int_0^\infty \exp(-DA) f(D)\, dD. \tag{4}$$

Cunningham [1990] provides an excellent review of different distributions that have been proposed in the literature for $f(D)$. Among these, the gamma distribution has met with considerable success (Okabe et al. [1972], Stapper [1973]). With this choice of $f(D)$, we arrive at the celebrated negative binomial yield model, whose parameters can be estimated from observed yield data. While these models and the corresponding estimation procedures are said to capture different degrees of *clustering* in the defect density, they do not directly take into account the spatial information contained in the wafer map data. As Stapper [1985] points out, "Modeling the distribution of the number of failing monitors per wafer...was therefore reduced to the search for a compounder that gave the correct distribution, mean, and variance."

Both of the two approaches outlined above can be combined to give the more general yield model

$$Y = Y_0 \int_0^\infty \exp(\, -DA\,)\, f(D)\, dD\,, \qquad (5)$$

which was in fact proposed early on by Murphy [1964]. As mentioned above, however, it is often difficult in practice to distinguish between the two types of clustering being accounted for in (5) solely on the basis of wafer map data. In regions where the defect density is high (say 20% more than the overall average D_0), random defects will be spatially clustered and hence may not be distinguishable from area-defects. (This is not a major concern in yield modeling where the predictive ability of a model is of primary importance with interpretation and conformance to reality being secondary.)

1.3 Overview

As noted at the outset, our primary interest is in process improvement through the routine identification of spatially clustered defects from wafer map data. Thus, our goal here is to estimate the *extent* of spatial clustering and the random defect density at the wafer map level. Therefore, throughout the rest of this paper, we will use a wafer-level definition of the parameters D_0 and Y_0.

Specifically, consider the two defect mechanisms described in Figure 1. We define Y_0 as the yield of the second wafer in Figure 1 with only clustered defects. Further, let π be the probability be that a randomly chosen chip in the first wafer (with only random defects) in Figure 1 is fault-free, and define D_0 through the expression

$$\pi = \exp(-D_0 A)\,. \qquad (6)$$

Hence, Y, the expected yield or probability that a randomly chosen chip is fault-free, is given by $Y = Y_0\pi = Y_0 \exp(-D_0 A)$, just as in expression (3). However, our definition of the parameters Y_0 and D_0 are made at the wafer-level and may not be strictly comparable to those in the yield modeling literature. We have also assumed a constant defect density. This is not a major concern in our framework as we allow D_0 to vary from wafer to wafer. Within a wafer, as previously noted, it is hard to separate clustered defects caused by large variation in the defect density from Y_0-defects. We do not feel it is necessary to a priori distinguish between different kinds of clustered defects and whether they are particle or process related failures. Once the spatially clustered defects are identified, the separation should be made by factory engineers on the basis of further analysis.

In our set-up, those chips in the observed wafer map in Figure 1 that belong to the cluster are said to comprise a large-area defect, while the remaining defective chips constitute small-area defects. By segmenting the failed circuits on an observed wafer map into small- and large-area defects defined in this way, we can estimate D_0 and Y_0 on a wafer-by-wafer basis. Estimation of D_0 and Y_0 has been considered by several authors in the literature, but most of them have taken a model-based approach (see, for example, Ham [1978], Paz and Lawson [1977], Warner [1981], and Stapper [1985]). In this paper, we consider a model-free approach that exploits the spatial patterns of defects in wafer map data from probe tests. We compare the performance of our method with the time-honored windowing technique employed by Seeds [1967], Paz and Lawson [1977], Warner [1981], and Mallory et al. [1983]. There have been other papers that have also tried to use spatial information to estimate D_0 and Y_0. In an earlier paper, Stapper [1976] attempts to extend the simple Poisson yield model by partitioning the wafer into regions in which the Poisson assumption might be more reasonable. Paz and Lawson [1977] and Saji of IBM (as reported in Stapper [1985]) also employ this partitioning scheme, but divide the wafer into different regions. Perhaps the de-clustering technique described by Stapper [1976] comes the closest in spirit to our method. Following his procedure, three or more adjacent defective chips are said to comprise a cluster. Once identified, these clusters are removed and the remaining defects are used to estimate the intensity of the random defects. Unfortunately, the behavior of procedures that employ a priori rules for determining a cluster will depend heavily on the yield of the wafer. Shortly, we will demonstrate that by conditioning on wafer yield, we avoid this problem and produce accurate estimates of both components of two defect mechanisms given in Figure 1.

The rest of this paper is organized as follows: the windowing method is reviewed and illustrated in Section 2; our method is described in Section 3; the two methods are compared via a simulation study in Section 4; and finally, in Section 5, we outline how the estimated quantities D_0 and Y_0 can be used for process improvement, both as metrics for routine process monitoring as well as measures of performance in designed experiments.

2. Estimating Y_0 and D_0 : The Windowing Method

By taking the logarithm of both sides of the expression (3), we find that

$$\log(Y) = \log(Y_0) - D_0 A, \tag{7}$$

where A is the active area of the circuit in question. Therefore, under this model, we expect to see a linear relationship between $\log(Y)$ and chip size. Because Y_0 and D_0 depend on process variation, however, it might be difficult to produce reliable estimates of these parameters from actual manufacturing data for products with different chip sizes. The windowing method seems to get around this problem by estimating Y_0 and D_0 on a wafer by wafer basis. The idea is quite simple: group neighboring chips together to form larger and larger *super-chips* and use the yield of these new circuits to estimate Y_0 and D_0 through the relationship given in (7). Here, we define a super-chip to be functional only if all of its component circuits are functional.

To make this description of the windowing procedure more precise, consider the final wafer map in Figure 1, and assume for simplicity that each circuit has an area of 1 square centimeter. In the second wafer map of Figure 2, we present one possible way to arrange (vertical) neighboring chips into super-chips of size 2 cm^2. Those super-chips for which at least one component circuit is defective are colored black, while the remaining super-chips are colored white. In a similar fashion, the third wafer map in this figure displays one possible way to group adjacent chips into super-chips of size 3 cm^2. The three maps in Figure 2 mark the points on the plot of the logarithm of super-chip yield versus super-chip area. Using equation (4), the estimates of Y_0 and D_0 obtained by the windowing technique are taken to be the intercept and slope, respectively, of a possibly weighted regression line fit to these three points. Not surprisingly, as the size of the super-chips increases, the variance in the accompanying yield estimate also increases. This fact motivates both our desire to keep the area of the super-chips small as well as the potential need for performing a weighted rather than an ordinary least-squares fit to the log-yield data.

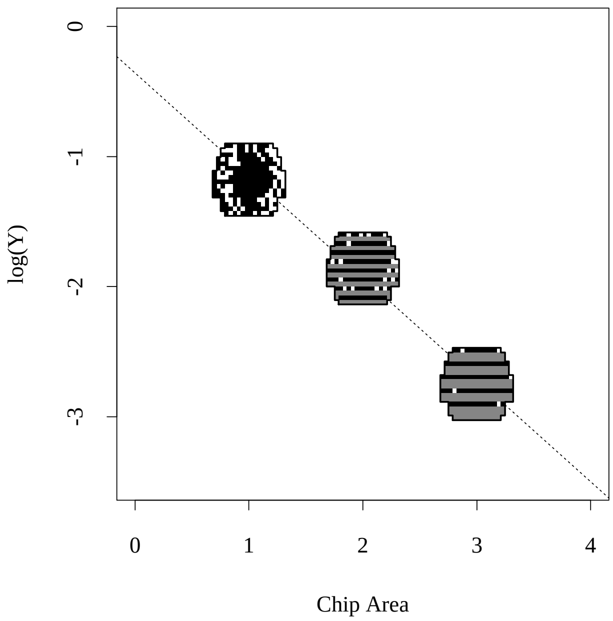

Figure 2. *A plot of the logarithm of super-chip yield versus area. The dashed line represents an ordinary least squares fit to these three points.*

Using the data from Figure 2, we arrive at an estimate of 0.68 for Y_0 and 0.78 defects/cm^2 for D_0. Referring back to Figure 1 and the definitions given in the previous section, the true value for Y_0, the yield of the second wafer in Figure 1, is 0.7. The functionality of each chip in the map of random defects in Figure 1 was decided by the toss of a fair coin, and hence the expected yield of this defect mechanism is 0.5. From equation (3) and our assumption that A is 1 square centimeter, we find that the true value of D_0 is 0.7 defects/cm^2. As part of the simulation study to be discussed in Section 4, we constructed 100 maps following this recipe. In the end, we found that the mean value of the estimates of Y_0 was 0.77 and the mean value of the estimates of D_0 was 0.79 defects/cm^2. This simulation suggests that there is some bias present in the estimates produced by the windowing procedure. Interestingly, this bias can be explained through the use of *join-count statistics* (see Moran [1948]), a critical component of our proposed method to be described in the next section.

3. Estimating Y_0 and D_0 by Analyzing Spatial Patterns

Our methodology for estimating Y_0 and D_0 first classifies the defective chips on an observed wafer map as either belonging to a large- or a small-area defect. This classification is based on a two step procedure of smoothing and then thresholding a given wafer map. Once classified, let N_l and N_s denote the number of defectives in each category, and let N denote the total number of chips on the wafer. Then, we can estimate Y_0 and π by the quantities

$$(N - N_l)/N \qquad \text{and} \qquad (N - N_l - N_s)/(N - N_l)$$

respectively. Finally, we arrive at an estimate for D_0 by transforming the estimate for π using equation (6).

3.1 Smoothed Wafer Maps

From each observed binary wafer map, we generate a smoothed map that is based on local averages (local defect probabilities). To be more precise, at each chip position, we consider a neighborhood around the given chip and calculate the proportion of the chip's neighbors that are defective. The choice of the neighborhood is flexible, although in practice we have typically employed 3×3 or 5×5 neighborhoods. Further, one could use a weighted average with some specified set of weights instead of a simple proportion. The computed proportions are then used to produce smoothed maps like the ones presented in Figure 3. Essentially, the proportions are transformed to values between zero and one by applying a sequence of transformations: a variance-stabilizing transformation followed by centering and scaling and finally a probability-integral transformation (see the appendix and Hansen et al. [1995a] for details). Figure 3 is a grey-scale map of these transformed values with grey-scales ranging between white (indicating a value of zero) and black (indicating a value of one). By construction, dark chips (values close to one) have too many defective neighbors relative to the overall yield of the wafer. Figure 3 presents four sets of observed and smoothed wafer maps; all four observed wafer maps have the same yield of 30% but the patterns of clustered defects are quite different.

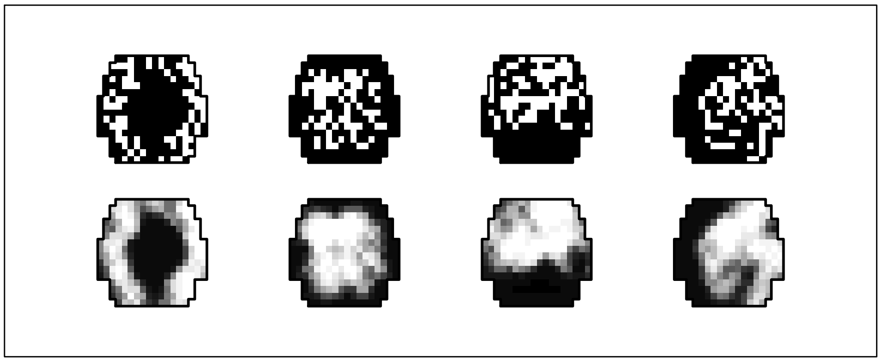

Figure 3. *Four examples of smoothed wafer maps. Each wafer has an overall yield of 30%.*

The size and location of possible large-area defects can now be estimated by choosing a threshold value for our smoothed maps and coloring black all defective chips with values above the threshold. The remaining chips will be colored white. The two extreme threshold values of one and zero have the following interpretations. A threshold value of one means that all the chips will be turned white, implying that all of the defective chips in the observed wafer map were generated randomly. On the other hand, a threshold value of zero implies that all the defective chips in the wafer map are part of a cluster. In the left-hand panel of Figure 4, we present the smoothed version of the observed wafer map appearing in Figure 1. The remaining three maps in this figure were obtained by thresholding this smoothed map with threshold values of 0.99, 0.9 and 0.5, respectively. By lowering the threshold value in this way, we create a sequence of nested clusters. As can be seen by comparing the results in Figure 4 with the original large-area defect in Figure 1, if the threshold is too large we underestimate the size of the cluster. On the other hand, a very low threshold leads to too many chips being included in the cluster and the resulting estimate becomes visually displeasing. In the next section, we will describe a simple rule for selecting the threshold value of our smoothed maps that yields accurate estimates of clusters, if they exist.

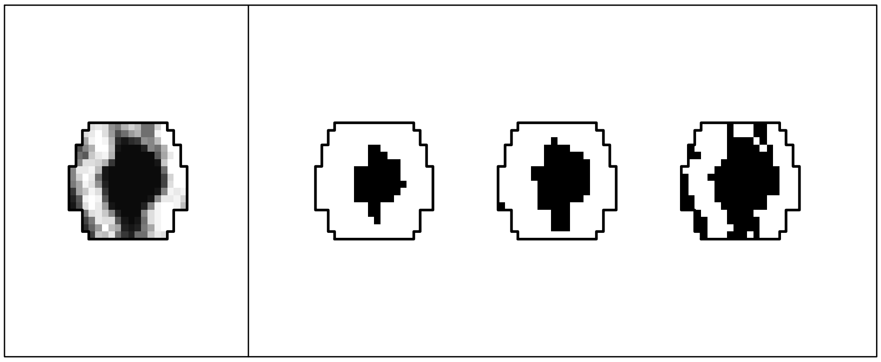

Figure 4. *Thresholding the smoothed map on the left at 0.99, 0.9 and 0.5.*

3.2 Choosing a Threshold

So far, we have constructed smooth wafer maps by transforming local defect probabilities, and have demonstrated that by properly thresholding these maps, we can produce reasonable estimates of large-area defects. In this section, we describe a technique for choosing threshold values based again on summary statistics derived from these local averages. Recall that under our framework as illustrated in Figure 1, defects can be classified into two categories: i) large-area defects, and ii) small area-defects distributed randomly across the wafer. Therefore, if we have properly identified the large-area defects, the defective chips in the remaining part of the observed wafer map should not exhibit any spatial patterns and should pass any test for spatial randomness. Thus, we begin with a threshold value of one which corresponds to declaring that the entire wafer consists of random defects and apply a test of spatial randomness. If this test fails, we reduce the threshold value, identify the corresponding area-defects and, conditional on these chips being defective, apply the test of spatial randomness to the remaining part of the wafer. The test proceeds in this way sequentially, until we reach a threshold value at which the hypothesis of complete spatial randomness is accepted. This value is chosen as the appropriate threshold value.

There are many different test for spatial randomness available in the literature (see, for example, Cressie [1991]). Many of these are based on the notion of *join-count* statistics (Moran [1948]). We have investigated the properties of these join-count statistics and have found a particular linear combination of two join-count statistics to perform well in our context. In addition, we have extended this test procedure to the conditional situation, where we need to test for spatial randomness in part of the wafer map after conditioning on the observed pattern in the remaining part. The details of this procedure are given in Hansen et al [1995b].

Returning to the present problem, Figure 5 illustrates how our strategy for choosing the threshold value works in practice. Consider again the observed wafer map in Figure 1. First, we choose a threshold of one for the smoothed version of this map, and recall that by doing so we are in fact declaring that there are no clusters present on the wafer. We apply our test for spatial randomness and compute the value of the $Z-$score for our test statistic. The y-coordinate of the center of the rightmost wafer map corresponds to this Z-score value. In addition, we have colored each defective chip grey in the accompanying wafer map. Since the Z-score value is large, we lower the threshold to 0.95 and identify chips that we will view as comprising a cluster. Conditional on this large-area defect, we apply the test of spatial randomness to the remaining part of the wafer. Corresponding to a threshold of 0.95 in Figure 5, we plot the value of the resulting Z-score, coloring the estimated large-area defect black and the remaining defective chips grey. Similarly, by considering threshold values of 0.9, 0.85 and 0.8, we complete the plot in Figure 5, which we refer to as a threshold profile.

Recall that we will choose our threshold to be the largest value that yields a large-area defect *consistent* with a given wafer map. Consistency in this case corresponds to choosing

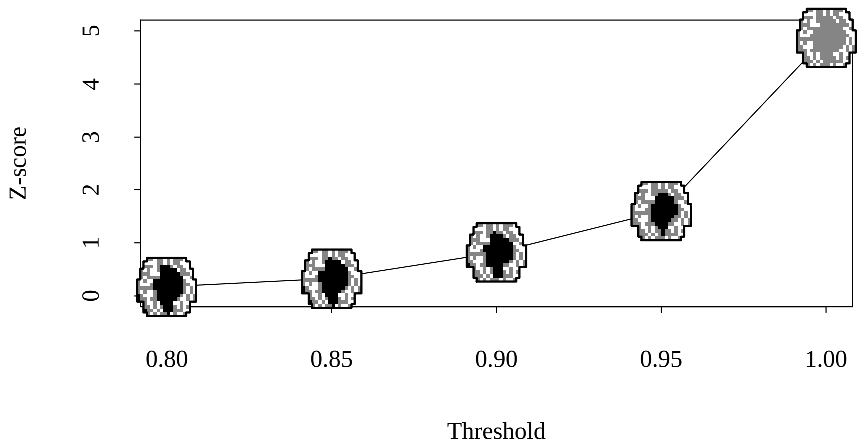

Figure 5. *Threshold profile for the final wafer map in Figure 1.*

an appropriate critical value for the sequence of Z-scores in the threshold profile. We have experimented with a number of different choices for this critical value. When the estimated large-area defects produced by this technique are to be examined visually, experience indicates that a high critical value should be chosen, say 1.5 or larger. In so doing, we are introducing a bias by intentionally underestimating the large-area defects to ease interpretation of the resulting clusters. On the other hand, empirical evidence suggests that a critical value between 0.5 and 0.7 is sufficiently low to overcome most of this bias and in fact yields remarkably accurate estimates of Y_0, the size of the underlying clusters. In the second row of Figure 6, we present the results of this procedure applied to the smoothed maps of Figure 3, using a critical value of 0.6. The original maps are presented in the first row for easy comparison. While we view the routine visual inspection of such thresholded maps as an important component in understanding the wafer fabrication process, in the present context, it is the size of these estimates that is interesting and not necessarily their configuration. Following the simple scheme outlined at the beginning of this section, we use these thresholded maps to derive estimates of D_0 and Y_0. For example, we arrive at an estimate of 0.71 for both D_0 and Y_0 for the first wafer in Figure 3 (the final wafer in Figure 1). As calculated in the previous section, the actual value for both D_0 and Y_0 was 0.7. As we will see, our procedure appears to be dramatically less biased than the windowing technique. In the next section, we describe a series of simulations designed to systematically compare these two procedures.

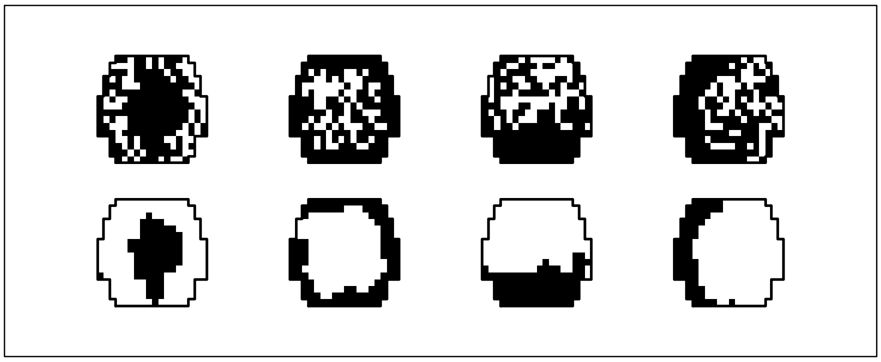

Figure 6. *Optimally chosen thresholds: 0.88, 0.74, 0.88 and 0.94, respectively, using a critical value of 0.6.*

4. A simulation study

4.1 Large-area defects

Recall the framework in Figure 1 where each observed wafer map is the superposition of two destructive forces. The first creates large-area defects, while the second generates defective chips uniformly across the wafer. In addition to the bull's eye pattern presented in Figure 1, we will consider three other cluster signatures. With varying intensities, these patterns are plotted in the top row of each table in Figure 7. By design, each set of four patterns corresponds to Y_0 values of 0.8, 0.7, 0.6, and 0.5, respectively. Each of these patterns is

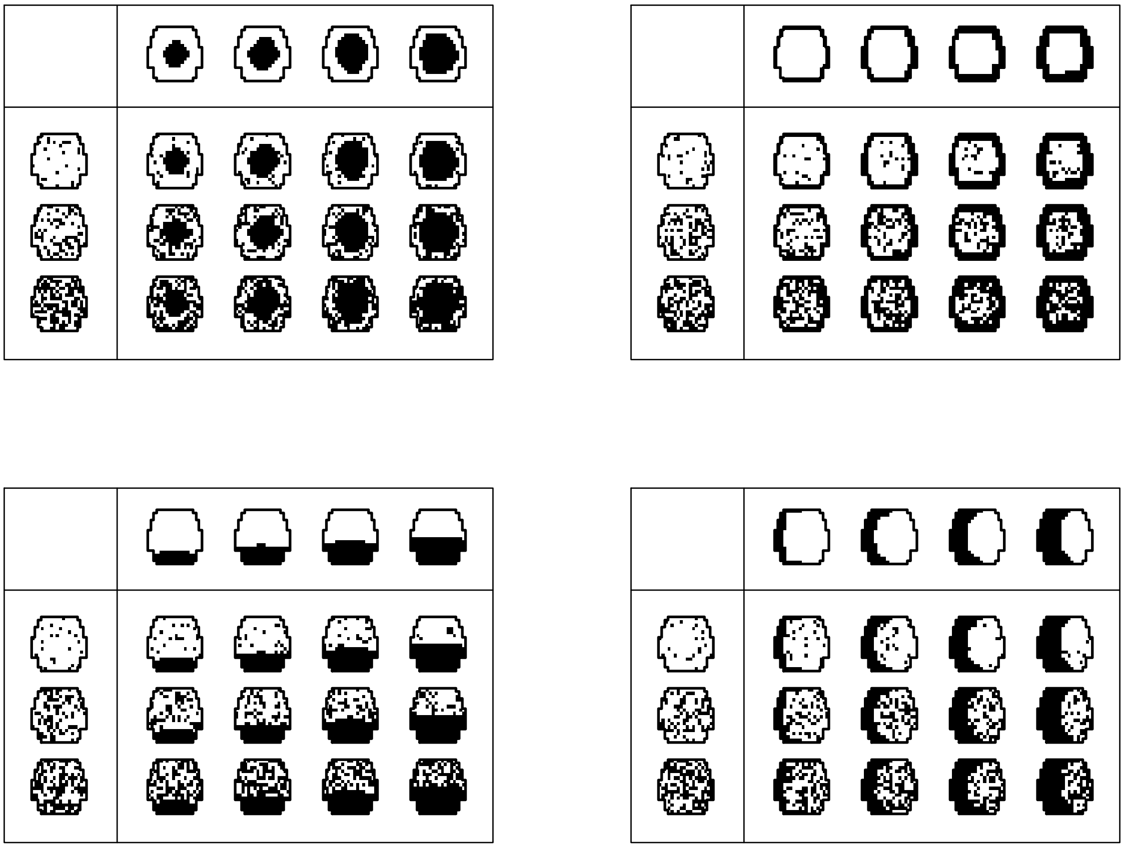

Figure 7. *Graphical description of our simulations.*

associated with a well-known manufacturing problem. The bull's eye defect pattern, for example, is often the result of a problem in the plasma area. The second type of cluster we consider causes the chips on the edge of the wafer to fail. This edge yield degradation is very common and is often traced to a lack of flatness on the wafer, handling problems, or design issues. Large-area defects that result in chip loss at the bottom of the wafer are the next pattern included in our simulations. This pattern can be the result of uneven heating during a diffusion process. The final clustered defect we consider causes chips to fail in a crescent moon pattern around one edge of the wafer. Based on previous experience, a fab engineer that notices this problem might decide to immediately look at the Rapid Thermal Anneal (RTA) process.

4.2 Small-area defects

In addition, we consider three different intensity levels for the spatially random defects. These are also shown in Figure 7. Along the leftmost row of each table we present binary maps like the first map in Figure 1 illustrating the random defect process with π taken to be 0.9, 0.7 and 0.5, respectively. Within each cell of these tables we combine the large- and small-area defects to generate an observed map similar to the final map in Figure 1.

4.3 The simulations

We generated 100 binary wafer maps corresponding to each entry in the tables of Figure 7. For each map, two estimates of Y_0 were calculated, one by windowing and one by our thresholding technique. Recall that the windowing technique depends on a regression of the logarithm of super-chip yield on the size of the super-chips. Therefore, if all of the super-chips of a given size are defective, the windowing estimate is not defined. Rather than attempt to impose ad hoc procedures for handling these cases, we simply dropped them from our analysis. Therefore, for many of the low yielding combinations of large- and small-area defects, a second set of 100 wafer maps had to be generated.

One component of these simulations not mentioned thus far has to do with the wafer maps themselves. The maps given in Figure 7 are each made up of 250 chips. In order to expand the applicability of our simulation results, we also considered maps having 100 and 500 chips per wafer. In each of these cases we constructed tables similar to those in Figure 7 and generated (possibly) two sets of 100 wafers per cell.

5. Simulation Results

5.1 Uniformly Distributed Defects

Suppose that each chip on a wafer is defective with probability π, independent of the status of the remaining chips. Under this simple model, defects are said to be distributed uniformly across the wafer and hence Y_0 is taken to be 1. We begin our simulation study by examining the behavior of the estimates of Y_0 in this context. For each of fourteen different values of π shown in Figure 8, we generated 250 binary wafer maps. Boxplots of the distributions for the two estimates of Y_0 are given in Figure 8. The results for the windowing technique should be viewed with caution. For low yielding wafers, there is high probability that this estimate is not defined. In fact, at a yield of 50%, we could only estimate Y_0 using this technique in 99 out of the 250 simulated wafers. Naturally the situation worsens as the yield decreases. Our method, however, works well in this simple situation across the full range of yields studied.

5.2 Area Defects

We now examine the performance of the estimates of Y_0 when genuine area defects are present. Recall that each of the four basic patterns were run at four levels of Y_0 (0.8, 0.7, 0.6, 0.5) and three levels of background noise (0.9, 0.7, 0.5) as described in Figure 7 of Section 4.1. In Figures 9a-c, we present simple plots of the average computed over 100 simulations of the window estimate of Y_0 versus that obtained by our thresholding procedure. Each point in these plots is represented by a wafer map. The color of the map indicates the associated value of π with the convention that the darker the color, the lower the value of π. In addition, within each plot, points corresponding to the same value of π are joined by a line. This should help the reader differentiate between the effect of varying Y_0 and that of varying π. Tables of these estimates and associated standard errors are given in the Appendix.

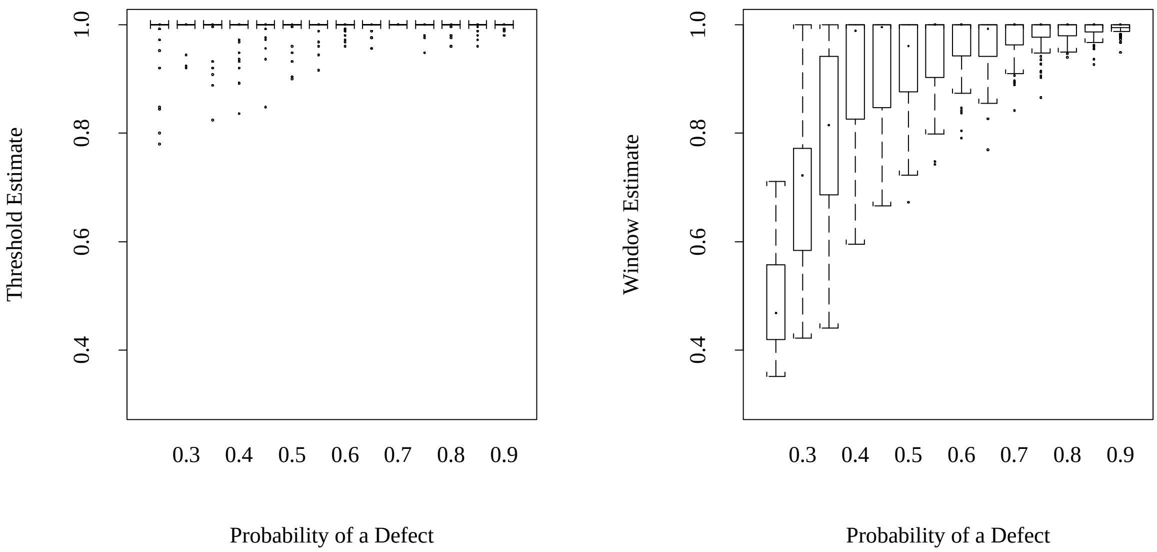

Figure 8. *The two estimates of Y_0 compared for several values of π.*

We observe immediately that the performance of each estimator improves considerably as we move from 100 to 500 chips per wafer. In addition, in wafers with smaller number of chips, windowing provides a more accurate estimate of the size of the smallest ring effect than our own procedure. Upon closer examination, we find that the thresholding technique has difficulty distinguishing between this model and a uniform defect distribution for low values of π. Given the classical notion of clustering as a failure of groups of contiguous chips, it is understandable that our technique would have trouble detecting the thinnest radial defects among our cluster signatures (see Figure 7). We can remove this effect entirely by employing neighborhoods that adapt to the shape of the wafer near the edges for constructing both the smoothed maps and the associated joint count statistics. A detailed discussion of this approach is available from the second author.

As hinted at in Section 2, both techniques are biased. From the standpoint of our procedure, this bias is worse for models with extremely low values of π (high D_0). Basically, as the yield of the random defect mechanism decreases, we incorrectly classify more and

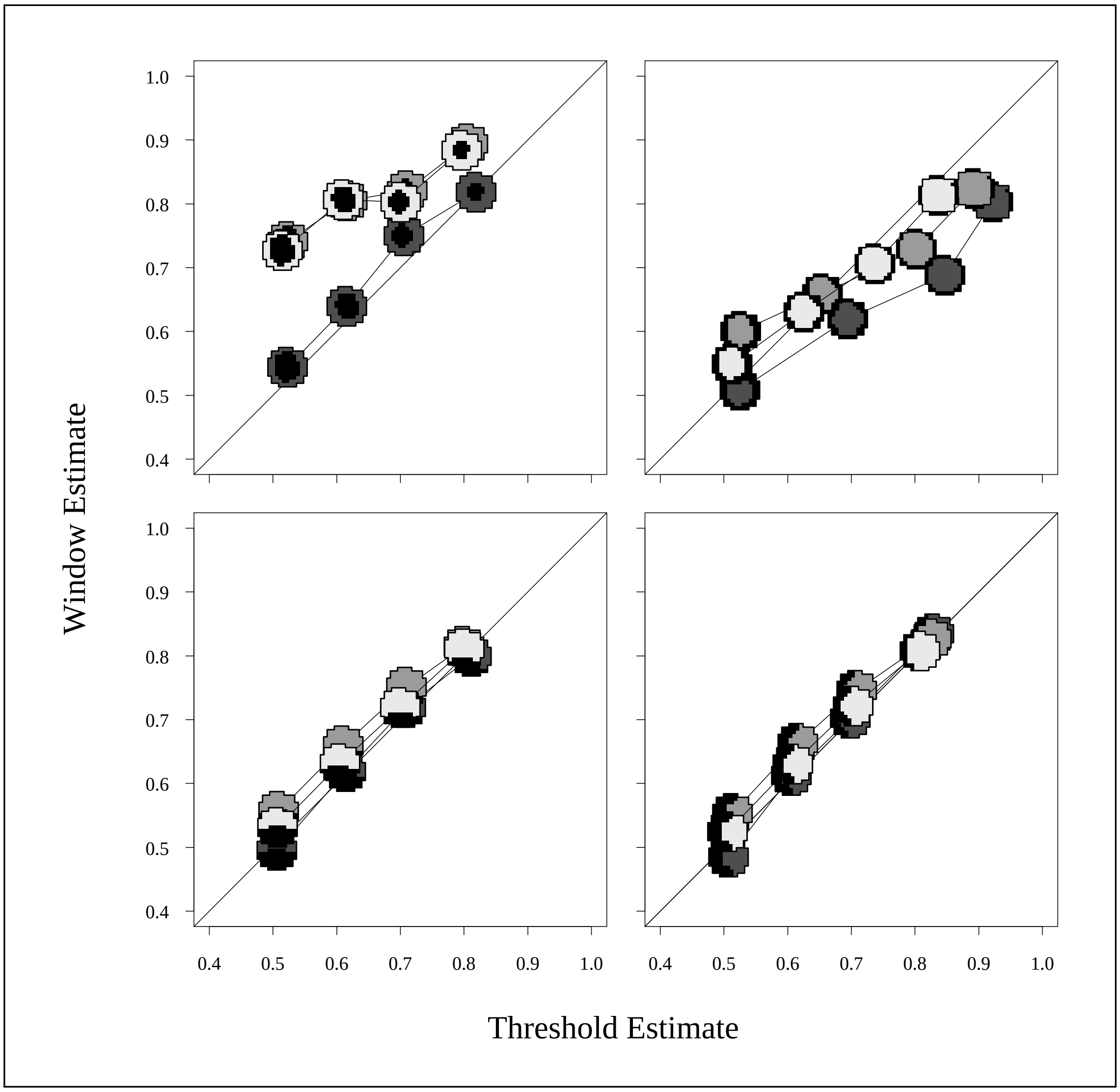

Figure 9a. *Comparison of window and threshold estimates, 100 chips per wafer.*

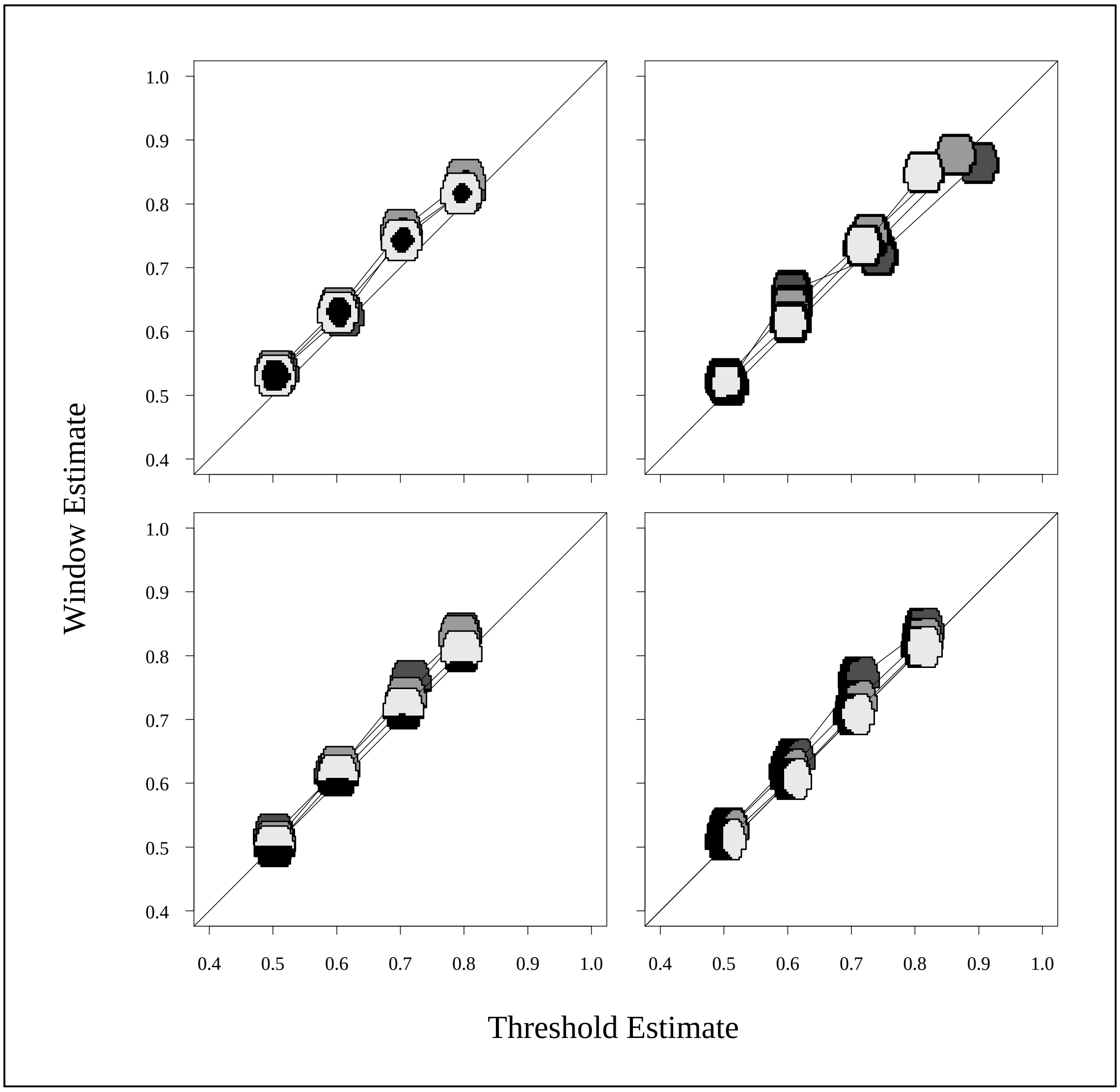

Figure 9b. *Comparison of window and threshold estimates, 250 chips per wafer.*

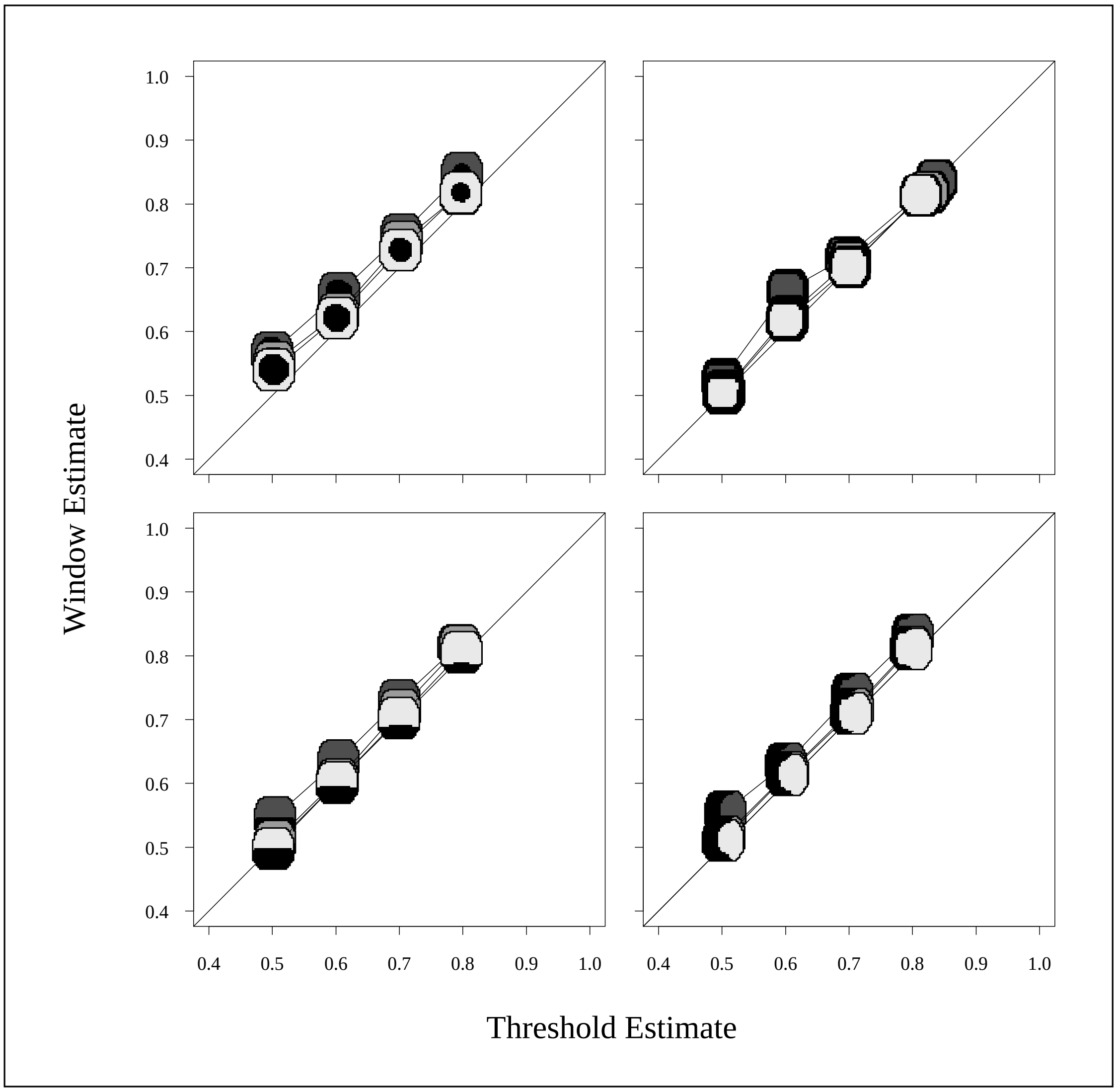

Figure 9c. *Comparison of window and threshold estimates, 500 chips per wafer.*

more chips from the edges of the cluster as belonging to the small-area defect. The bias in the windowing technique can be explained similarly, but because there is no direct spatial intuition behind this technique, the explanation is more difficult. Finally, we should comment on the somewhat paradoxical behavior of the observed bias in a few of the windowing estimates in Figures 9a-c. It would appear that in some cases, this estimate actually improves as π decreases. Recall, however, that for simplicity we have chosen to drop from our analysis those maps for which the window estimate was undefined, artificially reducing the bias in this procedure.

One interesting result of this analysis comes from the interplay between the breakdown point of each estimation technique. In the case of our thresholding procedure, failure corresponds to declaring that Y_0 is 1. This happens precisely when the binary wafer map is indistinguishable from one generated by a uniform defect density with an extremely low value of π. As noted above, the windowing method also tends to fail under under these low-yielding conditions. Perhaps not surprisingly, then, the two methods tend to fail together. For example, in those cases for which the bias of the windowing technique decreased with decreasing π, a corresponding decrease in the bias of our procedure is also observed, although the effect tends to be less significant.

In the tabular summary of these simulations in the appendix, we find that in addition to having less bias, the variance of our threshold estimate is much smaller than that of the window estimate. This can be seen in Figure 10 which presents four sets of smoothed histograms of the simulation results. These correspond to the simulations described by the entries in the first row and second column of the tables in Figure 7. In each, the solid line describes the behavior of the threshold estimate of Y_0 over the 100 simulations, while the dashed line describes the behavior of the window estimate over the same 100 wafer maps. The solid vertical line is at 0.7, the true value of Y_0 for these simulations.

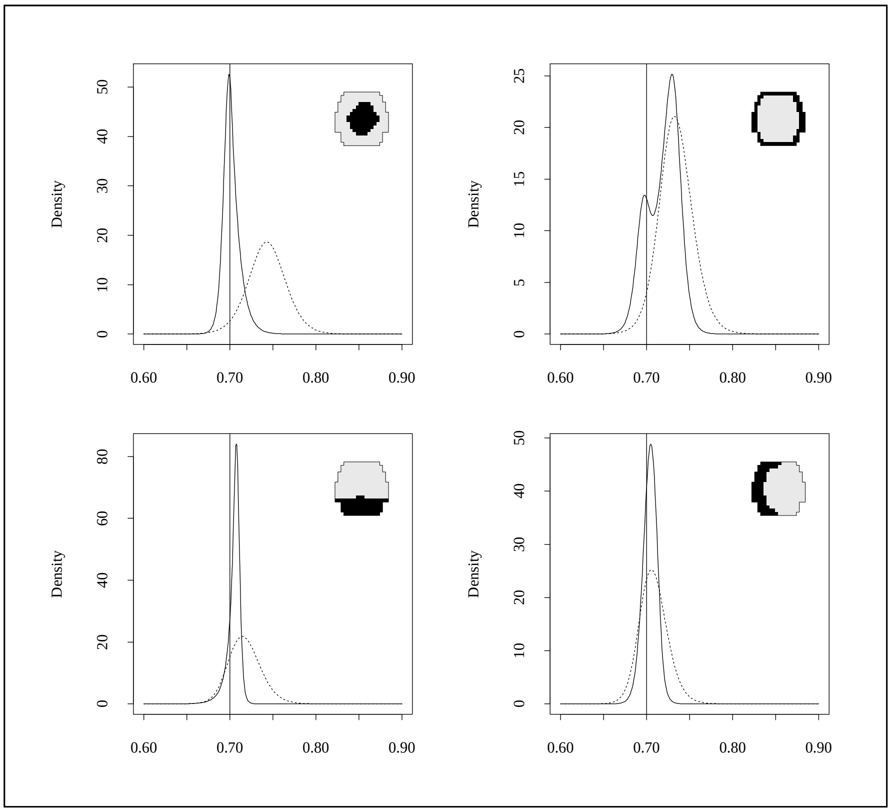

Figure 10. *Distribution of the estimates for various simulations.*

6. Concluding Remarks

We have described a model-free approach for estimating D_0 and Y_0 based on probe test at the wafer map level. The approach exploits the spatial information to obtain wafer-level estimates of defect clustering and random defect density. As shown through the simulation study, the method compares favorably with the windowing method currently available in the literature.

As noted earlier, our interest in estimating these quantities arose from a process improvement point of view. Although the impact of defect clustering on yield is widely recognized within the industry, currently no routine monitoring of factory data is done to detect the nature and extent of spatial clustering. The methods discussed in this paper will be useful in addressing this concern. For example, control charts based on the estimated values of D_0 and Y_0 can be developed and used to track changes in defect density and clustering over time. Given our definitions of the parameters D_0 and Y_0, such comparisons will be mostly meaningful within the same product type. The results in this paper will also be useful in analyzing data from designed experiments aimed at reducing defect clustering. Taam and Hamada [1993] discuss such experiments and use a variation of join-count statistics to measure the effectiveness of the design factor settings in reducing defect clustering. A more direct way to accomplish this is by estimating D_0 and Y_0 using the methods described in this paper and modeling these measures (or their logistic transforms) as a function of the design factor settings. These and other potential applications of our methodology will be developed in more detail in a forthcoming paper.

7. Acknowledgements

This work is a result of the joint efforts of the MSE Center of Excellence of the Engineering Research Center and the Statistical Research Center working in cooperation with AT&T Microelectronics facilities in Orlando, Reading, Allentown and Madrid. The authors would like to give special thanks to Dave Anderson, Dick Donovan, J. R. Hoyer, Daryl Pregibon, Bill Troutman and Steve Whitlock for their cooperation and support. V. N. Nair's research was supported in part by NSF Grant DMI-9501217 and by the Center for Display Technology and Manufacturing, University of Michigan.

8. References

[1] Cunningham, J.A. [1990] "The Use and Evaluation of Yield Models in Integrated Circuit Manufacturing," IEEE Trans. Semicond. Manufact. Vol. 3, N. 2, p. 60-71.

[2] Friedman, D.J, M. H. Hansen, and V.N. Nair [1995] " Process Improvement in Integrated Circuit Fabrication Through Spatial Analysis of wafer Map Data," (in preparation).

[3] Ham, W.E. [1978] "Yield-Area Analysis: Part I - A Diagnostic Tool for Fundamental Integrated Circuit Process Problems," RCA Review, Vol. 39, p. 231-249.

[4] Hansen, M.H., V.N. Nair, D.J. Friedman [1995a] "Identification of Spatial Patterns of Defect Clustering in Wafer Map Data," (in preparation).

[5] Hansen, M.H., V.N. Nair, D.J. Friedman [1995b] "Monitoring Wafer Map Data from Integrated Circuit Fabrication Processes for Spatially Clustered Defects," (in preparation).

[6] Lee, A. J. [1990]. "U-Statistics, Theory and Practice," Marcel Dekker, New York.

[7] Mallory, C.L., D.S. Perloff, T.F. Hasan, and R.M. Stanley [1983] "Spatial Yield Analysis in Integrated Circuit Manufacture," Solid State Technology, November, 1983, p. 121-127.

[8] Moran, P.A.P. [1948] "The Interpretation of Statistical Maps," J. Royal Stat. Soc. B, Vol. 10, p. 243-251.

[9] Murphy, B.T. [1964] "Cost-Size optima of monolithic integrated circuits," Proc. IEEE, Vol. 52, p. 1537-1545.

[10] Okabe, T., Nagata, M., and Shimada, S. [1972] "Analysis of yield of integrated circuits and a new expression for the yield," Electrical Engineering in Japan, Vol. 92, p. 135-141.

[11] Paz, O. and Lawson, T. [1977] "Modification of Poisson Statistics: Modeling Defects Induced by Diffusion," IEEE Journal of Solid-State Circuits, Vol. SC-12, p. 540-546.

[12] Seeds, R. B. [1967] "Yield and cost analysis of bipolar LSI," presented at IEEE International Electron Devices Meeting, Washington, D.C.

[13] Stapper, C.H. [1973] "Defect density distribution for LSI yield calculations," IEEE Trans. Electron. Devices, Vol. ED-20, p. 655-657.

[14] Stapper, C.H. [1976] "LSI Yield Modeling and Process Monitoring," IBM J. of Res. Develop., 20, 228-234.

[15] Stapper, C.H. [1985] "The effects of wafer to wafer defect density variations on integrated circuit defect and fault distributions," IBM J. of Res. Develop., 29, 1, 87-97.

[16] Taam, W. and M. Hamada [1992] "Detecting Spatial Effects from Factorial Experiments: An Application in Integrated- Circuit Manufacturing," To appear in Technometrics.

[17] Warner, R.M. [1981] "A note on IC-yield statistics," Solid- State Electronics, 24, 11, 1045-1047.

9. Appendix

We begin this appendix by introducing some notation. Let $\mathcal{W}$ denote a collection of chip sites that constitute a wafer map, and let W denote the total number of chips on the wafer. Throughout this discussion, we will use a script letter when referring to a set, and let the corresponding upper case letter denote its cardinality. Similarly, we will let $\mathcal{D} \subset \mathcal{W}$ denote the collection of defective chips on the wafer and D their number. Using this notation, the yield of the wafer is simply $1 - D/W$.

Technical Details: Smoothed Wafer Maps

As indicated above, our estimation procedure begins by generating smoothed wafer maps. The goal of our smoothing operation is to highlight regions with a large number of defective chips. As a visual process, this identification involves judging how closely the number of defective chips in small areas agree with what might be expected by considering the yield of the wafer as a whole. Toward this end, for each $i \in \mathcal{W}$, we let $\mathcal{N}_i$ denote the collection of chip locations that we take to be the neighbors of chip i and let N_i denote their number. To each chip in $j \in \mathcal{N}_i$, we assign a weight $w_i(j)$ and we require that the sum of the $w_i(\cdot)$ over the neighbors of chip i be equal to one.

Next, for each $i \in \mathcal{W}$, we define a local kernel estimator of the proportion of defective chips on the wafer

$$\hat{p}_i \;=\; \frac{1}{N_i} \sum_{j \in \mathcal{N}_i} I_{\mathcal{D}}(j),$$

where for each $j \in \mathcal{W}$, $I_{\mathcal{D}}(j)$ is the indicator function of the event that the chip at position i is defective. The situation where the defective chips are distributed randomly across the wafer will be referred to as the null hypothesis of *complete spatial randomness* (CSR). Under this hypothesis, the variables $I_{\mathcal{D}}(j)$, are independent and identically distributed as Bernoulli random variables with constant probability $p_j = p$, $j \in \mathcal{W}$. By conditioning on the yield of the wafer, we can set $p = D/W$. The presence of clusters of defective chips will result both in areas in which $\hat{p}_j$ consistently larger than D/W and in areas for which $\hat{p}_j$ consistently smaller than D/W.

Kernel smoothers of this type are particularly attractive in this application because they lend themselves easily to a number of distributional calculations. If we condition on the yield of the wafer, then under the hypothesis of complete spatial randomness the number of defective chips occurring in a neighborhood $\mathcal{N}_i$ of a given chips i has a hypergeometric distribution with parameters N, D and W. Of course, we could, in principal, write down the distribution for $\hat{p}_i$ for each $i \in \mathcal{W}$. However, if N is small compared to W, we can safely employ a binomial approximation to this distribution, and if $w_i(j) = 1/N$ for all $j \in \mathcal{N}$, the marginal distributions of the $\hat{p}_i$ follow immediately.

Keeping with the simple case of $w_i(j) = 1/N$ for the moment, we observe that our $\hat{p}_i$ will have different variances under the hypothesis of complete spatial randomness for different overall yields, further complicating the interpretation of the variation in these smoothed maps. To remove this dependence on yield we define a new map comprised in which over each position i, we plot $\hat{s}_i$, where

$$\hat{s}_i \;\; = \;\; \sin^{-1} \sqrt{\hat{p}_i},$$

the arcsine-square root transformation. It is well know that the $\hat{s}_i$ will have an approximately normal distribution with mean y and variance $1/(4N_i)$. More generally, if we allow the $w_i(\cdot)$ to define an arbitrary kernel over $\mathcal{N}_i$, then the $\hat{s}_i$ will have an approximately normal distribution with mean y and variance $(\sum [w_i(j)]^2)/4$. Although the variance of the $\hat{s}_i$ no longer depends on y, we still have to consider the impact of chips with incomplete neighborhoods on the overall variation in this map.

To get a feel for the magnitude of this effect, in the wafers presented above, in the first wafer map of Figure 1, 20% of the chips have incomplete three-by-three neighborhoods, while 40% have incomplete five-by-five neighborhoods. To get around this problem, we will standardize our $\hat{s}_i$, so that under the hypothesis of complete spatial randomness, at each chip position we have an observation from the standard normal distribution. We identify areas of clustered defects by evaluating at each point of the wafer map the probability that the standardized $\hat{s}_i$ would be as small or smaller than its observed value. Essentially, then, we simply transform the entire map using the standard normal distribution function. Following Diggle (1984), we refer to these plots as probability surfaces.

Technical Details: Join-Count Statistics

In implementing our thresholding technique, we have made considerable use of a test for spatial randomness based on the join-count statistics:

$$T_{\mathcal{D}} \;=\; W^{-1} \sum_{i \in \mathcal{W}} \sum_{j \in \mathcal{W}} w_i(j) I_{\mathcal{D}}(i) I_{\mathcal{D}}(j) \qquad (A.1a)$$

and

$$T_{\mathcal{W}\setminus\mathcal{D}} \;=\; W^{-1} \sum_{i \in \mathcal{W}} \sum_{j \in \mathcal{W}} w_i(j) I_{\mathcal{W}\setminus\mathcal{D}}(i) I_{\mathcal{W}\setminus\mathcal{D}}(j)\,. \qquad (A.1b)$$

The main reason for this choice is that it is relatively easy to study the properties of these statistics as we can compute their first two moments exactly under null and alternative hypotheses of interest. These statistics were first proposed by Moran (1948) and have been discussed in great detail by Cliff and Ord (1981). See also Taam and Hamada (1993).

After a little algebra, we find that under the hypothesis of complete spatial randomness the means of $T_{\mathcal{D}}$ and $T_{\mathcal{W}\setminus\mathcal{D}}$ conditional on $\hat{p}$ are

$$\mathbf{E}\left(T_{\mathcal{D}} \mid \hat{p} = \frac{D}{W}\right) \;=\; \frac{D\,[D-1]}{W\,[W-1]}$$

and

$$\mathbf{E}\left(T_{\mathcal{W}\setminus\mathcal{D}} \mid \hat{p} = \frac{D}{W}\right) \;=\; \frac{[W-D]\,[W-D-1]}{W\,[W-1]}\,,$$

assuming that $w_j(j)$ is not zero for all $j \in \mathcal{W}$. If $w_j(j) > 0$ for at least one chip j, then a simple correction factor is added to the above expression that takes into account the fact that $I_{\mathcal{D}}(j)^2 = I_{\mathcal{D}}(j)$.

The variance-covariance matrix of the pair $T_{\mathcal{D}}, T_{\mathcal{W}\setminus\mathcal{D}}$ depends not only on D and W, but also on the arrangement of chips on the wafer $\mathcal{W}$ and the weights $\{w_j(\cdot)\}$. We can express the second moment of $T_{\mathcal{D}}$, for example, as

$$\mathbf{E}\left(T_{\mathcal{D}}^2 \mid \hat{p} = \frac{D}{W}\right) = \alpha_1 \frac{D}{W} + \alpha_2 \frac{D\,[D-1]}{W\,[W-1]} + \alpha_3 \frac{D\,[D-1]\,[D-2]}{W\,[W-1]\,[W-2]} + \alpha_4 \frac{D\,[D-1]\,[D-2]\,[D-3]}{W\,[W-1]\,[W-2]\,[W-3]},$$

where $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$, and α_1 is zero if $w_j(j)$ is zero for all $j \in \mathcal{W}$. Roughly speaking, these expressions are derived by considering the proportion of terms in the expansion of the second moment that involve two, three and four unique indicators, respectively. Finally, assuming that $w_j(j)$ is zero for all $j \in \mathcal{W}$, the covariance between $T_{\mathcal{D}}$ and $T_{\mathcal{W}\setminus\mathcal{D}}$ can be found using

$$\mathbf{E}\left(T_{\mathcal{D}} T_{\mathcal{W}\setminus\mathcal{D}} \mid \hat{p} = \frac{D}{W}\right) = \alpha_4 \frac{D\,[D-1]\,[W-D]\,[W-D-1]}{W\,[W-1]\,[W-2]\,[W-3]},$$

where α_4 is the same as the quantity appearing in the expression for the second moments of $T_{\mathcal{D}}$ and $T_{\mathcal{W}\setminus\mathcal{D}}$. Again, if for some j, $w_j(j)$ is not zero, then lower-order products would be included in this expression.

To get the exact distribution of these test statistics, we have to resort to the permutation distribution based on all possible allocations of D defects among the W different chips in the wafer $\mathcal{W}$. While it is not difficult to generate wafer maps that follow this distribution, it would be computationally intensive to depend on these samples for approximate critical values for our test statistics. As we have seen, the higher moments of $T_{\mathcal{D}}$ and $T_{\mathcal{W}\setminus\mathcal{D}}$ depend on the specifics of the wafer map. This would mean that new samples would have to be drawn for each different map in production. For this reason, we investigated the adequacy of using large-sample normal approximations to the distribution of $T_{\mathcal{W}\setminus\mathcal{D}}$ and $T_{\mathcal{D}}$. Using the theory of $U-$statistics (see for example Lee, 1990), it can be shown that the conditional distribution of the test statistics, suitably normalized, have a limiting bivariate normal distribution as $W \to \infty$. This fact can be used to obtain approximate critical values for the test statistics.

Finally, we consider a *clustered* alternative hypothesis where we condition on the value of $I_{\mathcal{D}}(j)$ for a particular region of the wafer. To be more precise, let $\mathcal{C} \subset \mathcal{W}$ denote a collection of chip sites that we take to be a cluster. Now, consider the distribution of the pair $(T_{\mathcal{D}}, T_{\mathcal{W}\setminus\mathcal{D}})$ conditional on the event that the chips in $\mathcal{C}$ are defective. It is possible to compute the means and variances of $T_{\mathcal{D}}$ and $T_{\mathcal{W}\setminus\mathcal{D}}$ under this alternative.

$$\begin{aligned} \mathbf{E}\left(T_{\mathcal{D}} \mid \hat{p} = \frac{D}{W},\, \mathcal{C}\right) &= T(\mathcal{C},\mathcal{C}) + [T(\mathcal{C},\mathcal{C}^c) + T(\mathcal{C}^c,\mathcal{C})]\,\frac{D-C}{W-C} \\ &\quad + T(\mathcal{C}^c,\mathcal{C}^c)\,\frac{[D-C]\,[D-C-1]}{[W-C]\,[W-C-1]}, \end{aligned}$$

and

$$\mathbf{E}\left(T_{\mathcal{W}\setminus\mathcal{D}} \mid \hat{p} = \frac{D}{W},\, \mathcal{C}\right) = T(\mathcal{C}^c,\mathcal{C}^c)\,\frac{[W-D]\,[W-D-1]}{[W-C]\,[W-C-1]},$$

where for any sets $\mathcal{I}, \mathcal{J} \subset \mathcal{W}$, we define

$$T(\mathcal{I},\mathcal{J}) = W^{-1} \sum_{i \in \mathcal{W}} \sum_{j \in \mathcal{W}} w_i(j) I_{\mathcal{I}}(i) I_{\mathcal{J}}(j)\,.$$

Observe that $T(\mathcal{D},\mathcal{D})$ is simply $T_{\mathcal{D}}$. The expressions for the other moments are somewhat complicated and are not presented here. In general, these quantities require the calculation of constants that depend on the configuration of the wafer map as well as the shape of the cluster $\mathcal{C}$. It is not difficult to establish that, under this set up, assuming that $W - C$ is large, then $T_{\mathcal{D}}$ and $T_{\mathcal{W}\setminus\mathcal{D}}$ can be approximated by a bivariate normal distribution with the above moments.

Tables of Results

		Y_0			
		0.8	0.7	0.6	0.5
	0.9	0.798	0.701	0.609	0.513
		(0.014)	(0.016)	(0.020)	(0.020)
D_0	0.7	0.802	0.711	0.621	0.524
		(0.044)	(0.051)	(0.044)	(0.044)
	0.5	0.821	0.714	0.611	0.527
		(0.080)	(0.075)	(0.071)	(0.074)

		Y_0			
		0.8	0.7	0.6	0.5
	0.9	0.839	0.735	0.620	0.513
		(0.033)	(0.041)	(0.028)	(0.018)
D_0	0.7	0.893	0.807	0.661	0.524
		(0.074)	(0.099)	(0.054)	(0.050)
	0.5	0.918	0.852	0.707	0.529
		(0.095)	(0.117)	(0.125)	(0.090)

		Y_0			
		0.8	0.7	0.6	0.5
	0.9	0.799	0.700	0.605	0.507
		(0.010)	(0.015)	(0.016)	(0.014)
D_0	0.7	0.799	0.712	0.608	0.519
		(0.031)	(0.030)	(0.039)	(0.034)
	0.5	0.817	0.699	0.611	0.507
		(0.077)	(0.057)	(0.058)	(0.055)

		Y_0			
		0.8	0.7	0.6	0.5
	0.9	0.808	0.703	0.610	0.508
		(0.016)	(0.012)	(0.015)	(0.014)
D_0	0.7	0.826	0.711	0.619	0.512
		(0.047)	(0.042)	(0.037)	(0.034)
	0.5	0.819	0.694	0.604	0.505
		(0.079)	(0.070)	(0.060)	(0.058)

Table 1a. *Threshold estimates of Y_0 and their standard errors, 100 chips per wafer. Tables are referenced to Figure 7.*

		Y_0			
		0.8	0.7	0.6	0.5
	0.9	0.892	0.806	0.797	0.717
		(0.041)	(0.048)	(0.053)	(0.080)
D_0	0.7	0.891	0.816	0.817	0.742
		(0.091)	(0.124)	(0.114)	(0.158)
	0.5	0.815	0.763	0.640	0.546
		(0.139)	(0.170)	(0.148)	(0.106)

		Y_0			
		0.8	0.7	0.6	0.5
	0.9	0.815	0.706	0.624	0.548
		(0.031)	(0.033)	(0.028)	(0.031)
D_0	0.7	0.820	0.738	0.663	0.586
		(0.093)	(0.118)	(0.103)	(0.120)
	0.5	0.811	0.694	0.626	0.499
		(0.146)	(0.138)	(0.134)	(0.117)

		Y_0			
		0.8	0.7	0.6	0.5
	0.9	0.812	0.719	0.628	0.529
		(0.028)	(0.027)	(0.028)	(0.025)
D_0	0.7	0.811	0.759	0.661	0.547
		(0.089)	(0.099)	(0.117)	(0.098)
	0.5	0.781	0.706	0.614	0.494
		(0.159)	(0.147)	(0.140)	(0.121)

		Y_0			
		0.8	0.7	0.6	0.5
	0.9	0.809	0.725	0.637	0.524
		(0.028)	(0.033)	(0.035)	(0.026)
D_0	0.7	0.830	0.756	0.674	0.544
		(0.089)	(0.114)	(0.123)	(0.098)
	0.5	0.826	0.707	0.612	0.482
		(0.144)	(0.163)	(0.128)	(0.115)

Table 1b. *Window estimates of Y_0 and their standard errors, 100 chips per wafer. Tables are referenced to Figure 7.*

D_0	Y_0 = 0.8	0.7	0.6	0.5
0.9	0.795	0.702	0.602	0.503
	(0.005)	(0.010)	(0.010)	(0.011)
0.7	0.802	0.701	0.603	0.504
	(0.019)	(0.026)	(0.026)	(0.024)
0.5	0.800	0.703	0.610	0.510
	(0.039)	(0.045)	(0.049)	(0.046)

D_0	Y_0 = 0.8	0.7	0.6	0.5
0.9	0.814	0.719	0.605	0.504
	(0.019)	(0.018)	(0.010)	(0.010)
0.7	0.864	0.730	0.607	0.502
	(0.058)	(0.031)	(0.023)	(0.019)
0.5	0.901	0.745	0.606	0.509
	(0.067)	(0.073)	(0.045)	(0.032)

D_0	Y_0 = 0.8	0.7	0.6	0.5
0.9	0.796	0.705	0.602	0.502
	(0.004)	(0.008)	(0.007)	(0.007)
0.7	0.793	0.709	0.604	0.503
	(0.015)	(0.018)	(0.019)	(0.020)
0.5	0.794	0.716	0.597	0.502
	(0.038)	(0.033)	(0.034)	(0.038)

D_0	Y_0 = 0.8	0.7	0.6	0.5
0.9	0.810	0.704	0.605	0.503
	(0.010)	(0.009)	(0.007)	(0.009)
0.7	0.812	0.709	0.603	0.506
	(0.021)	(0.020)	(0.022)	(0.021)
0.5	0.814	0.712	0.610	0.509
	(0.051)	(0.031)	(0.033)	(0.032)

Table 2a. *Threshold estimates of Y_0 and their standard errors, 250 chips per wafer. Tables are referenced to Figure 7.*

D_0	Y_0 = 0.8	0.7	0.6	0.5
0.9	0.816	0.743	0.629	0.531
	(0.019)	(0.023)	(0.020)	(0.021)
0.7	0.837	0.759	0.637	0.538
	(0.082)	(0.080)	(0.081)	(0.066)
0.5	0.821	0.755	0.632	0.541
	(0.144)	(0.168)	(0.143)	(0.132)

D_0	Y_0 = 0.8	0.7	0.6	0.5
0.9	0.849	0.734	0.615	0.520
	(0.024)	(0.020)	(0.016)	(0.019)
0.7	0.877	0.751	0.640	0.526
	(0.084)	(0.080)	(0.081)	(0.075)
0.5	0.865	0.721	0.658	0.512
	(0.132)	(0.154)	(0.145)	(0.108)

D_0	Y_0 = 0.8	0.7	0.6	0.5
0.9	0.807	0.717	0.612	0.501
	(0.017)	(0.019)	(0.016)	(0.015)
0.7	0.831	0.734	0.626	0.509
	(0.076)	(0.077)	(0.083)	(0.062)
0.5	0.831	0.755	0.615	0.517
	(0.136)	(0.155)	(0.141)	(0.120)

D_0	Y_0 = 0.8	0.7	0.6	0.5
0.9	0.814	0.708	0.607	0.512
	(0.016)	(0.017)	(0.014)	(0.015)
0.7	0.826	0.729	0.622	0.527
	(0.067)	(0.075)	(0.088)	(0.074)
0.5	0.843	0.765	0.635	0.529
	(0.137)	(0.152)	(0.151)	(0.117)

Table 2b. *Window estimates of Y_0 and their standard errors, 250 chips per wafer. Tables are referenced to Figure 7.*

		Y_0			
		0.8	0.7	0.6	0.5
	0.9	0.796	0.701	0.601	0.503
		(0.003)	(0.005)	(0.006)	(0.007)
D_0	0.7	0.797	0.704	0.603	0.503
		(0.013)	(0.014)	(0.016)	(0.015)
	0.5	0.798	0.703	0.605	0.499
		(0.025)	(0.032)	(0.034)	(0.027)

		Y_0			
		0.8	0.7	0.6	0.5
	0.9	0.813	0.701	0.602	0.502
		(0.014)	(0.006)	(0.005)	(0.005)
D_0	0.7	0.824	0.701	0.603	0.503
		(0.022)	(0.014)	(0.012)	(0.014)
	0.5	0.837	0.695	0.603	0.501
		(0.049)	(0.030)	(0.028)	(0.024)

		Y_0			
		0.8	0.7	0.6	0.5
	0.9	0.798	0.699	0.601	0.501
		(0.004)	(0.003)	(0.004)	(0.005)
D_0	0.7	0.796	0.701	0.603	0.504
		(0.010)	(0.014)	(0.013)	(0.010)
	0.5	0.793	0.700	0.604	0.504
		(0.026)	(0.029)	(0.027)	(0.024)

		Y_0			
		0.8	0.7	0.6	0.5
	0.9	0.798	0.703	0.602	0.501
		(0.004)	(0.004)	(0.005)	(0.007)
D_0	0.7	0.797	0.706	0.604	0.503
		(0.010)	(0.013)	(0.009)	(0.014)
	0.5	0.800	0.705	0.601	0.505
		(0.024)	(0.026)	(0.031)	(0.025)

Table 3a. *Threshold estimates of Y_0 and their standard errors, 500 chips per wafer. Tables are referenced to Figure 7.*

	Y_0			
	0.8	0.7	0.6	0.5
0.9	0.818	0.728	0.621	0.540
	(0.011)	(0.011)	(0.010)	(0.011)
0.7	0.817	0.741	0.628	0.551
	(0.042)	(0.042)	(0.057)	(0.055)
0.5	0.848	0.752	0.660	0.567
	(0.095)	(0.117)	(0.126)	(0.127)

	Y_0			
	0.8	0.7	0.6	0.5
0.9	0.814	0.701	0.618	0.504
	(0.010)	(0.008)	(0.011)	(0.006)
0.7	0.819	0.711	0.624	0.507
	(0.047)	(0.038)	(0.039)	(0.029)
0.5	0.836	0.716	0.665	0.525
	(0.100)	(0.107)	(0.130)	(0.099)

	Y_0			
	0.8	0.7	0.6	0.5
0.9	0.806	0.702	0.601	0.498
	(0.009)	(0.009)	(0.009)	(0.008)
0.7	0.815	0.715	0.606	0.510
	(0.041)	(0.039)	(0.031)	(0.031)
0.5	0.816	0.730	0.636	0.547
	(0.098)	(0.114)	(0.126)	(0.151)

	Y_0			
	0.8	0.7	0.6	0.5
0.9	0.811	0.710	0.613	0.511
	(0.011)	(0.010)	(0.011)	(0.009)
0.7	0.814	0.717	0.617	0.517
	(0.036)	(0.036)	(0.031)	(0.034)
0.5	0.833	0.740	0.631	0.556
	(0.092)	(0.098)	(0.119)	(0.132)

Table 3b. *Window estimates of Y_0 and their standard errors, 500 chips per wafer. Tables are referenced to Figure 7.*