Model-Free Estimation of Defect Clustering in Integrated Circuit Fabrication

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

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

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

David James
Statistics Research
Bell Laboratories
Murray Hill, New Jersey

Abstract

This paper describes a model-free method for estimating the size, shape and location of large-area defects, or clusters of defective chips, evident in binary probe test data. Unlike previous methods discussed in the literature, our approach makes extensive use of the location of failing chips within the wafer map to directly estimate clusters. An important byproduct of this analysis is a decomposition of wafer yield that attributes defective chips to either large- or small-area defects. Simulation studies show that our procedure is superior to the time-honored windowing technique for achieving a similar breakdown. In addition, by directly estimating defect clusters, we can provide engineers with a greater understanding of the manufacturing process. It has been our experience that routine identification of the spatial signatures of clustered defects and associated root-cause analysis is a cost-effective approach to yield and process improvement.

1. Introduction

Over the past 30 years, researchers have constructed various mathematical descriptions of wafer yield. Most of these models were developed to characterize the defect distribution as a function of device size and complexity, so that reliable capacity and cost estimates could be calculated. The early yield models were based on the Poisson distribution, but these have turned out to be inadequate, underestimating the yield for large chips (Cunningham [1990]). Poor predictions of this type stemmed from the fact that defect densities are subject to considerable spatial and temporal variations, effects that have been broadly described as defect clustering. Much of the subsequent work has been aimed at developing alternative distributions that better predict the number of defects either occurring within a single chip of a given area or across an entire wafer, in essence fitting a marginal distribution to counts of defects or failing devices. In general, while these models try to account for the extra variation in wafer yield caused by defect clustering, they do not directly exploit any of the spatial information contained in mapped wafer data. Unlike previous work in this area, we consider yield modeling primarily as a tool for process and quality improvement. Rather than view defect clustering solely through its impact on yield, we make explicit use of the spatial distribution of defective chips in probe data to identify potential process problems.

Throughout the paper, we consider binary wafer map data, where each chip (integrated circuit) is classified as either good or defective. We introduce a statistical procedure that routinely examines the patterns of failed chips on a given wafer map to identify the size, shape and location of clustered defects (should they exist), and flags wafers or lots that exhibit significant spatial clustering. We stress that this is quite different from previous modeling efforts that focus on marginal yield behavior to signal problems. It has been our experience that routine identification and root-cause analysis of the spatial patterns of clustered defects is a cost-effective approach to yield improvement. In the pages that follow we describe how this approach can be used as the basis for more informative process monitoring.

To illustrate the need for such a procedure, consider the plot in Figure 1a in which

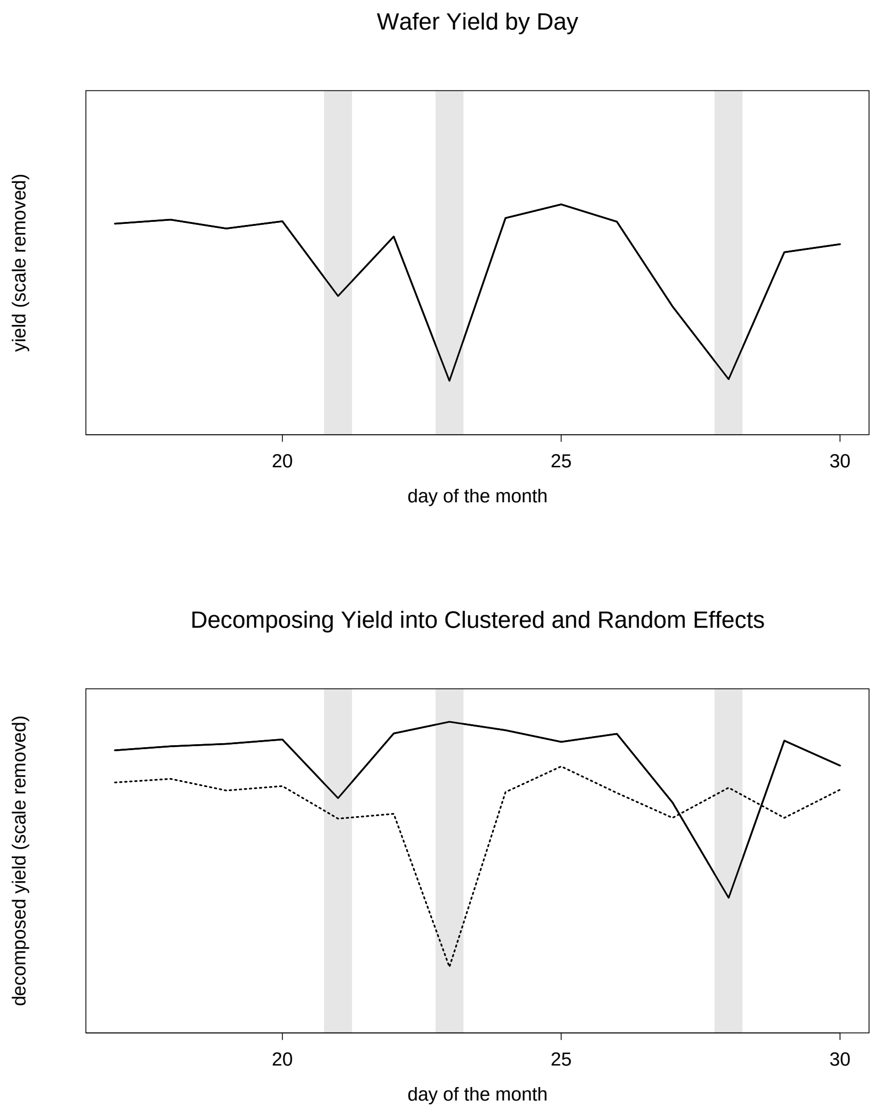

Figure 1. *Separating the two defect mechanism: (a) A simple plot of the average yield of lots probed during two weeks of production, and (b) attributing this yield loss to large- and small-area defect mechanisms (the solid line tracks the π_l component of yield while the dashed line follows π_s).*

the average lot-level yield of two weeks of production for a particular code is plotted as a function of time. In Figure 2a, we present three wafers selected from the lots probed on the $21st$, $23rd$ and $28th$ days of the month under study. While these wafers have similar yields, the arrangement of defective chips, the spatial defect patterns, are much different. Since spatially clustered defects are often due to process problems, information about their size, shape and location can be extremely valuable to engineers when identifying potential problems in their line. In contrast to the wafers in the first and third rows of this figure, the samples taken from the $23rd$ day of production seem to have a rather high, but uniform defect density. This density is much higher than the background density that appears in addition to the clustered defects in wafers from the other two days. While a much less common problem, a low uniform, defect density can be characterized completely by its overall level or yield.

In Figure 1b, we have separated the overall lot-level yield loss in Figure 1a into two components: a) loss from large-area clustered defects (typically attributable to systematic problems), and b) loss due to random or small-area defects. This segmentation is done on a wafer-by-wafer basis, and the results are illustrated in Figure 2b for our sample of wafers described above. It is interesting to note that except for the $23rd$ day of production, the proportion of random defects is fairly constant, while the proportion of chips lost to clustered defects varies substantially. Because our procedure for segmenting yield involves estimating the actual clustered defect, we can produce cluster signatures on a daily basis to summarize for engineers the types of yield problems that have been seen at the end of the line.

The purpose of this paper is to describe a model-free statistical approach for estimating defect clusters from binary probe test data. The method is model-free in the sense that we do not assume any particular distribution for the size, location or shapes of the clusters. Instead, we employ nonparametric smoothing techniques in concert with tests of spatial randomness to determine the clusters for each wafer map. It is, of course, possible to formulate spatial models for the large-area clusters if a manufacturing line is fairly stable

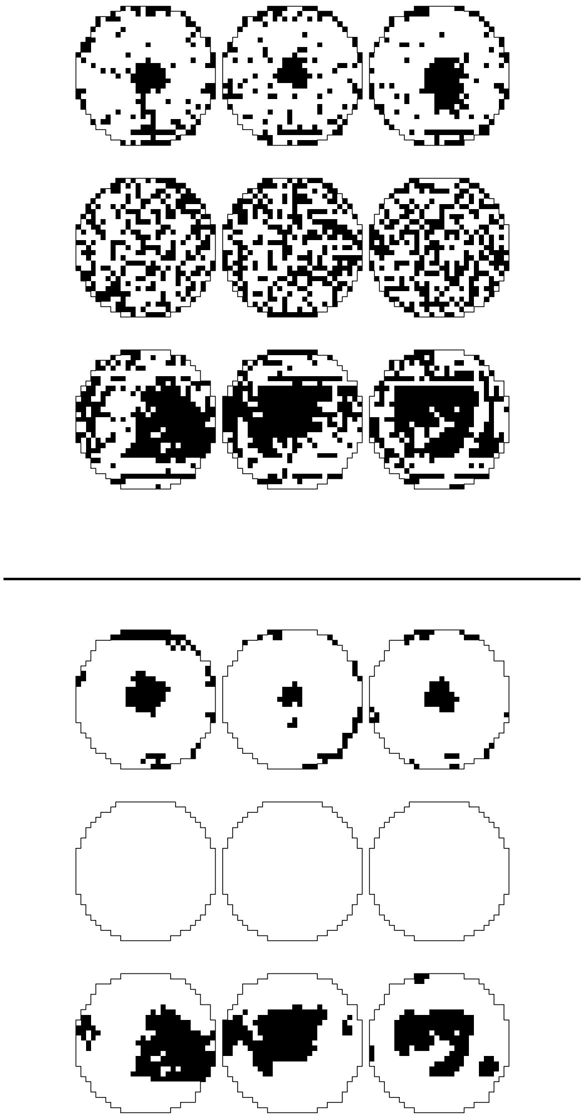

Figure 2. *Separating the two defect mechanism: (a) a sample of the binary probe data for lots processed on the 21st, 23rd and 28th days of the month and (b) the identified clusters.*

over time. Unfortunately, the data we have seen to date is too variable to allow for refined modeling. Our approach is geared to such an environment where information about clustering within individual lots can be used to identify machines or processing steps that are not meeting specifications. Understanding the variation in yield is the starting point for our process and quality improvement efforts, and this understanding demands some type of visual inspection of the wafer map data.

In the next section, we introduce the basic framework underlying our segmentation procedure. We also include a brief review of the traditional yield models in this section and compare them with our approach. In Section 3 we adapt the so-called windowing technique to estimate the large- and small-area yield loss plotted in Figure 1b on a wafer-by-wafer basis. The details of our approach are described in Section 4. A critical component in our analysis is that we have an estimate of the large-area defects, if they are present, on a wafer-by-wafer basis. In the final two sections, we compare our technique to the windowing procedure through a number of simulations and demonstrate that in addition to the advantage of producing wafer-level estimates of clusters, our procedure is much more stable.

2. A spatial model for wafer yield

2.1 Motivation

As noted in the introduction, 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 a chip with a given design, then under the Poisson model the probability that a chip has k defects is

$$P(k) \;=\; (D_0A)^k \exp(-D_0A)/k!\,,$$

independent of the defect counts for the remaining chips on the wafer. Therefore, such a chip is fault-free with probability

$$\pi \;=\; \exp(-D_0A)\,, \qquad (1)$$

again independent of the other chips on the wafer. Since this probability applies uniformly across the wafer, the resulting spatial model for the distribution of defective chips is referred to as a homogeneous Bernoulli process. While this model is not able to capture the variations we have seen in our manufacturing facilities, we will use it as a building-block and approximate the spatial distribution of defective chips across wafers by finite mixtures of spatially homogeneous Bernoulli processes. In this case, we view a wafer as being divided into distinct regions with a roughly constant yield for each region. Formally, this implies that the probability that a chip is fault-free $\pi = \pi(c)$ depends on c, the chip location. This approach is similar to the one taken by Yanagawa [1972], Gupta et al. [1974], Paz and Lawson [1977] and Saji of IBM (as reported in Stapper [1985]) who considered an extension of the Poisson model by partitioning each wafer into *fixed* regions where the Poisson assumption might be more reasonable. In general, however, we will allow the regions, and hence the values of $\pi(c)$, to vary from wafer to wafer and from lot to lot, depending on both the type of device (its technology and code) and the status of the manufacturing line.

To simplify matters, we have found that it is sufficient to consider only the special case of Bernoulli processes with two components: one corresponding to a region where the probability that a chip is fault-free $\pi(c)$ is low, and another where $\pi(c)$ is (relatively) high. The former corresponds to a cluster or large-area defect, while the latter is populated by random or small-area defects. In Figure 3 we present a graphical description of the resulting two-component Bernoulli model: $\pi(c)$ is essentially zero for chips in the black region of the leftmost wafer, while the middle wafer exhibits random defects corresponding to a constant probability $\pi(c) = \pi = 0.5$. The binary wafer map we see at probe, the rightmost wafer in Figure 3, is in fact a superposition of these two components. Because it is difficult to provide a formal probability specification that adequately describes all of the cluster shapes and sizes that we encounter in practice, our approach will be to condition on a given wafer map and use the spatial information about the arrangement of failing chips to identify clustered defects through data-analytic techniques. In practice, we can only separate these regions when the two underlying intensities are quite different, in which case we can even

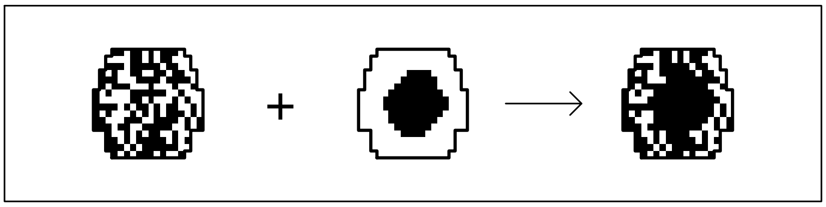

Figure 3. *An idealized view of a two component Bernoulli process.*

tolerate a certain degree of inhomogeneity in the probabilities $\pi(c)$ within each region.

To justify this simplification and motivate its usefulness for process improvement, we return again to the factory data presented in Figures 1 and 2. Clusters like the ones presented in our first and third samples 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. Process-related clusters are often caused by one or more process step not meeting the required specifications. For example, an uneven growth of oxide could cause all the chips in a certain sector of the wafer to fail. In each of these cases, knowing the the shape and location of the clusters can give process and product engineers valuable clues to identify the underlying cause. By routinely monitoring wafer map data for spatial clustering of this type, we can more readily identify potential problems in the fabrication process and thereby improve yield.

Spatially random defects, on the other hand, 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 process 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. Unlike clustered defects which have a specific size and shape, the spatially random component of an observed wafer map can be completely characterized by the proportion of chips that were fault-free in this region. While we expect less variability from these random defect levels, it is important to routinely monitor both components of wafer yield for irregularities. The second sample in Figure 1 serves to further illustrate this point.

2.2 Mathematical details

Consider again the two-component, spatial Bernoulli model introduced above. Let π_1 denote the probability that a chip in region $\mathcal{C}_1$ is fault-free, with similar definitions for π_2 and $\mathcal{C}_2$. We will use the symbol $\mathcal{W}$ to refer to the collection of chip locations for a particular layout so that $\mathcal{W} = \mathcal{C}_1 \cup \mathcal{C}_2$. Throughout this discussion, the division of $\mathcal{W}$ into $\mathcal{C}_1$ and $\mathcal{C}_2$ is treated as random, having a spatial distribution that depends on the type of chip being fabricated and a set of manufacturing conditions. For simplicity, we will assume that the probabilities π_1 and π_2 are fixed given the same set of conditions. (It is trivial to extend our treatment to the case in which these quantities are also random: If we let f_i denote the underlying distribution on these probabilities, then in what follows π_i can be thought of as the mean of f_i.) For $c \in \mathcal{W}$,

$$\begin{aligned} P(c \text{ is good}) &= P(c \in \mathcal{C}_1)\, P(c \text{ is good} | c \in \mathcal{C}_1) + P(c \in \mathcal{C}_2)\, P(c \text{ is good} | c \in \mathcal{C}_2) \\ &= P(c \in \mathcal{C}_1)\, \pi_1 + P(c \in \mathcal{C}_2)\, \pi_2 \end{aligned}$$

Suppose that $\mathcal{W}$ contains W chips, and let Y denote the yield of a wafer manufactured under a given set of conditions,

$$Y = \frac{1}{W} \sum_{c \in \mathcal{W}} I_{\{c \text{ is good}\}}\,,$$

where $I_{\{c \text{ is good}\}}$ is a binary random variable that takes on the value one if c is good and zero otherwise. The expected yield of a wafer manufactured under the specified conditions is given by

$$EY = \frac{1}{W} \sum_{c \in \mathcal{W}} P(c \text{ is good})\,,$$

and hence

$$\begin{aligned} EY &= \frac{1}{W}\sum_{c\in\mathcal{W}} P\left(c\in\mathcal{C}_1\right)\pi_1 + \frac{1}{W}\sum_{c\in\mathcal{W}} P\left(c\in\mathcal{C}_2\right)\pi_2 \\ &= \pi_1\frac{EC_1}{W} + \pi_2\frac{EC_2}{W}\,, \end{aligned}$$

where C_i represents the number of chips that intersect the region $\mathcal{C}_i$,

$$C_i \;=\; \sum_{c\in\mathcal{W}} I_{\{c \text{ intersects } \mathcal{C}_i\}}\,,$$

and EC_i is the expected size or number of chips contained in the area. In this case $I_{\{c \text{ intersects } \mathcal{C}_i\}}$ is again a binary random variable that is one if c touches $\mathcal{C}_i$ and zero otherwise.

If $\pi_2 \approx 0$, then

$$EY \;=\; \pi_1\frac{EC_1}{W} \;=\; \pi_1\left(1-\frac{EC_2}{W}\right)\,, \tag{2}$$

since $\mathcal{W} = \mathcal{C}_1 \cup \mathcal{C}_2$. Written in this way, π_1 corresponds to the small-scale or random defect mechanism and $1 - EC_2/W$ is the yield associated with large-scale defects or clusters. For the remainder of this paper, we will only consider this two component spatial Bernoulli model and we will set

$$\pi_s \;=\; \pi_1 \qquad \text{and} \qquad \pi_l \;=\; 1-\frac{EC_2}{W}\,,$$

representing the small and large-area defect components, respectively. We will also drop the subscript corresponding to region number and set $\mathcal{C} = \mathcal{C}_2$, an area of low (essentially zero) yield. To be clear, both π_s and the spatial distribution of $\mathcal{C}$ are assumed to depend on the type of device in question as well as processing conditions in the line, and hence can vary over time. Because of this, we choose to classify defective chips on each wafer as being part of either a large- or small-area defect. If a manufacturing line is stable over a given period of time, we can combine the separate estimates of π_s and $\mathcal{C}$ (or π_l) to construct more refined spatial yield models. In any case, monitoring each yield component and examining lots with significant spatial clusters is an effective tool for process improvement.

2.3 Relationship to traditional yield models

In motivating the decomposition (2), we began by considering mixtures of spatially homogeneous Bernoulli processes, noting that this model can also be obtained by assuming that each wafer is divided into regions where a simple Poisson model holds. In this case, if we associate each region with a different defect density, then as the size and shape of the regions vary, we in effect put a distribution f on a *random* defect density that we will denote by D. Under the resulting compound Poisson yield model, we have that

$$EY \;=\; \int_0^\infty \exp(\,-DA\,)\, f(D)\, dD\,. \tag{3}$$

Cunningham [1990] provides an excellent review of different distributions for $f(D)$ that have been proposed in the literature. Among these, the gamma family 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 easily estimated from observed yield data, summarized at the wafer level. 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.” Koren, Koren and Stapper [1994] perform a similar type of search, but try to find a correct defect distribution by matching the empirical distribution of defect counts per chip to mixtures of various mathematical models, including the Poisson and compound Poisson.

To achieve a better fit to manufacturing data, a so-called “gross yield factor” Y_0 is usually incorporated in the yield model (3) so that

$$EY \;=\; Y_0 \int_0^\infty \exp(\,-DA\,)\, f(D)\, dD\,. \tag{4}$$

In this setting, what is meant precisely by Y_0 varies considerably in the literature. Following Murphy [1964], some engineers view a gross yield loss to be one which is independent of chip size and one in which every chip within the area is defective. Given this interpretation, Y_0 is viewed as a proxy for process failures, while the remaining component in (4) becomes a surrogate for particle based problems. Ham [1978] takes this even farther, viewing Y_0 as an "area usage factor" representing consistent yield loss at the edge of the wafer to handling problems and increased misalignment. Paz and Lawson [1977], on the other hand, take a more liberal view of Y_0, demonstrating that regions with high defect densities can also be modeled by Y_0. Unfortunately, it is virtually impossible to make these distinctions solely on the basis of wafer map data. Many fab engineers believe that to understand the nature of an observed fault, some micro-inspection is required, making a strict particle- versus process-related failure separation an extremely time consuming proposition. We believe, however, that this characterization is not entirely relevant from the standpoint of monitoring the wafer map data collected at final probe: Consistent particulate failures that occur with sufficient intensity to produce large clusters of defective chips should send a strong signal that there is a problem in the process.

If f in (4) is taken to be a degenerate distribution with a single point mass at D_0, then the expression for expected chip yield becomes

$$EY \;=\; Y_0 \exp(\,-D_0 A\,), \tag{5}$$

a model originally proposed by Murphy [1964]. Here, as in equation (1), A represents the active or effective area of the device. Recall from the discussion at the beginning of this section that the popular negative binomial yield model is obtained by assuming that the density f in (4) is a member of the gamma family. In this case,

$$EY \;=\; Y_0 \left(1 + \frac{\bar{\lambda}}{\alpha}\right)^{-\alpha}, \tag{6}$$

where α is a cluster parameter and $\bar{\lambda}$ is the average number of defects per chip. To estimate the parameters in models like (5) and (6), Seeds [1967], Paz and Lawson [1977], Warner [1981], Hemmert [1981], Mallory et al. [1983] and Stapper [1989] all discuss a technique known as "windowing" that can be applied to individual wafer maps like those presented in Figure 1. Unlike the search for compounders mentioned above, windowing does take into account some spatial information about the placement of defective chips within a wafer. As a practical matter, when applied to probe data this procedure is blind to the subtle distinction between a heavy particulate event generating an area of low yield and more traditional process problems. Still, taking the liberal view of Y_0 discussed by Paz and Lawson [1977], the widowing technique provides a useful decomposition of wafer yield.

Our own breakdown of yield (2) into losses attributable to large-area clusters and random or small-area defects is roughly equivalent to fitting Murphy's model (5) wafer-by-wafer. The parameters π_l and π_s correspond to Y_0 and $\exp(-D_0A)$, respectively, with the added assumption that these quantities will vary from wafer to wafer and from lot to lot, and that their distributions depend on conditions in the production line. In the next section, we will illustrate a version of the windowing technique that can be used to produce these wafer-level estimates. In contrast to this approach and its *indirect* use of spatial information on failing devices, we have developed a procedure that is aimed directly at determining the size and shape of clustered defects on a wafer-by-wafer basis. In addition to producing dramatically more stable estimates, our new procedure allows engineers to develop a better understanding of spatial cluster process at work destroying chips in the line, ultimately leading to faster response times to identified problems and improved yields. Again we cite Figures 1 and 2 as examples of our point. 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 prespecified 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 3.

3. Estimating π_s and π_l : The windowing method

By taking the logarithm of both sides of the expression (5), we find that

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

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 3, and assume for simplicity that each circuit has an area of 1 square centimeter. In the second wafer map of Figure 4, 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 4 mark the points on the plot of the logarithm of super-chip yield versus super-chip area. Using equation (5), 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.

Using the data from Figure 4, we arrive at an estimate of 0.68 for Y_0 and 0.78 defects/cm^2

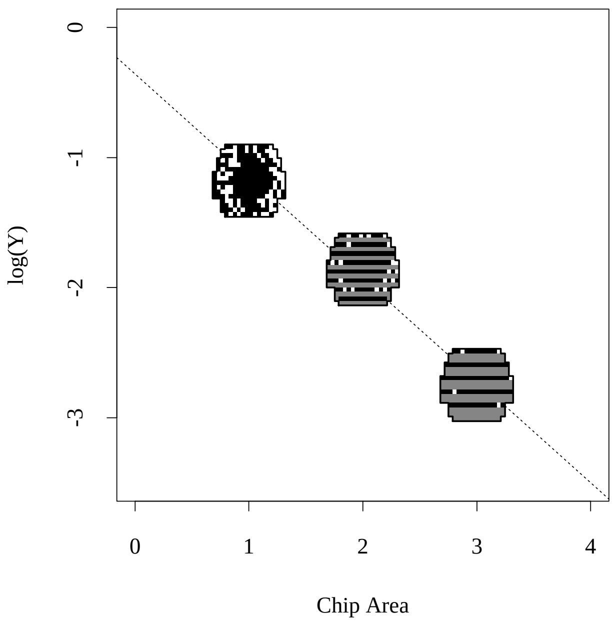

Figure 4. *A plot of the logarithm of super-chip yield versus area. The dashed line represents an ordinary least squares fit to these three points.*

for D_0. Referring back to Figure 3 and the definitions given in the previous section, the true value for Y_0, the yield of the second wafer in Figure 3, is 0.7. The functionality of each chip in the map of random defects in Figure 3 was decided by the toss of a fair coin, and hence the expected yield of this defect mechanism is 0.5. From equation (5) 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 5, 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.

We end this section by casting the classical windowing procedure in terms of our decomposition given in (2). Suppose we group neighboring chips into super-chips c^* of size k. Recall that a super-chip is fault-free only if all the original chips it contains are fault-free. Therefore, if a super-chip is located entirely within the region of random defects, the chance that it is good is π_s^k. Furthermore, according to our model in (2) we now have an expected yield of $\pi_s^k \pi_l$ where

$$\pi_l \;=\; 1 - \frac{EC}{W}\,, \qquad C \;=\; \sum_{c^* \in \mathcal{W}^*} I_{\{c^* \text{ intersects } c\}}\,,$$

and we have used the symbol $\mathcal{W}^*$ to refer to the collection of super-chips c^*. When the original map $\mathcal{W}$ contains a sufficiently large number of chips, we can ignore edge effects in defining π_l and view it as essentially constant for small values of k. In this case we can again take logarithms to find that for super-chips of size k,

$$\log(\,EY\,) \;=\; \log(\,\pi_l\,) + k\log(\,\pi_s\,)\,.$$

Therefore, even without directly appealing to the notion of active areas, we can apply the windowing procedure to derive wafer-level estimates of π_l and π_s, the yield losses associated from the clustered and random defect mechanisms, respectively.

4. Estimating π_s and π_l by analyzing spatial patterns

Our methodology for estimating π_s and π_l 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 W denote the total number of chips on the wafer. Then, we can estimate π_s and π_l by the quantities

$$\hat{\pi}_s = (W - N_l - N_s)/(W - N_l) \qquad \text{and} \qquad \hat{\pi}_l = (W - N_l)/W.$$

4.1 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 5. 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 for details). Figure 5 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 5 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.

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

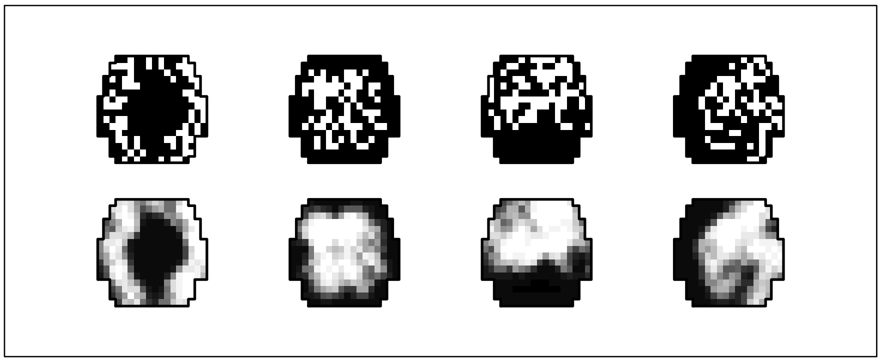

Figure 5. *Four examples of smoothed wafer maps. Each wafer has an overall yield of 30%.*

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 6, we present the smoothed version of the observed wafer map appearing in Figure 3. 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 6 with the original large-area defect in Figure 3, 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.

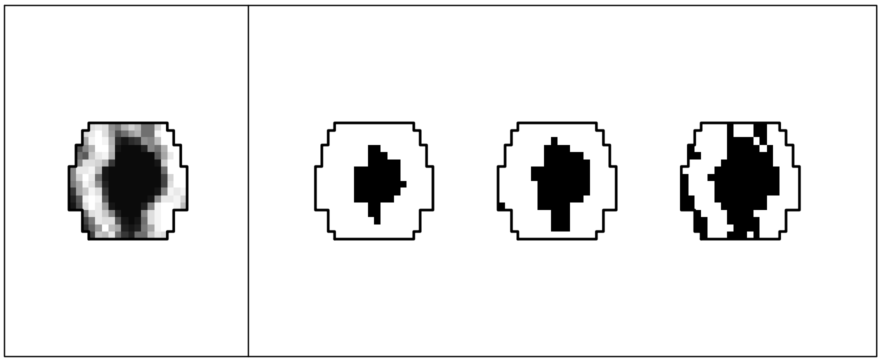

Figure 6. *Thresholding the smoothed map on the left at 0.99, 0.9 and 0.5.*

4.2 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 3, 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 testing 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 so-called *join-count* statistics (Moran [1948]). We have investigated the properties of these join-count statistics and have found that a particular linear combination of two such statistics to performs 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 [1996].

Returning to the present problem, Figure 7 illustrates how our strategy for choosing the threshold value works in practice. Consider again the observed wafer map in Figure 3. 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. (For complete details, the reader is referred to Hansen et al. [1996].) 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 7, 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 7, 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 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,

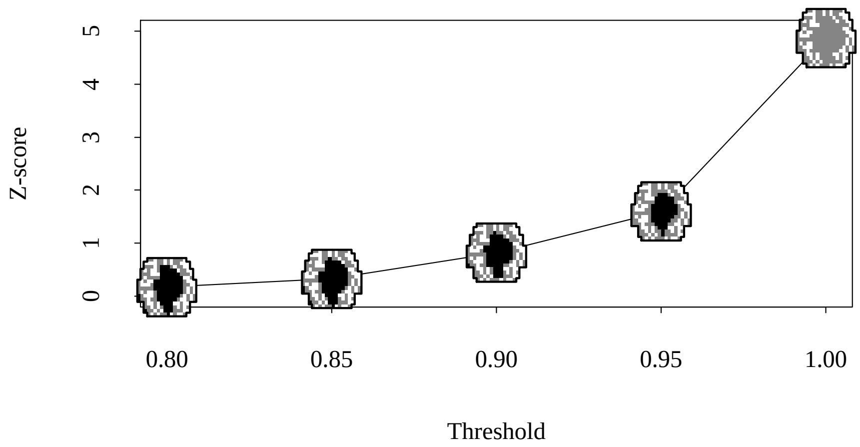

Figure 7. *Threshold profile for the final wafer map in Figure 3.*

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 π_l, the size of the underlying clusters. In the second row of Figure 8, we present the results of this procedure applied to the smoothed maps of Figure 5, 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 π_s and π_l. For example, we arrive at an estimate of 0.71 for both π_s and π_l for the first wafer in Figure 5 (the final wafer in Figure 3). As calculated in the previous section, the actual value for both π_l and π_s was 0.7. As we will see, our procedure appears to be dramatically less biased than the windowing technique.

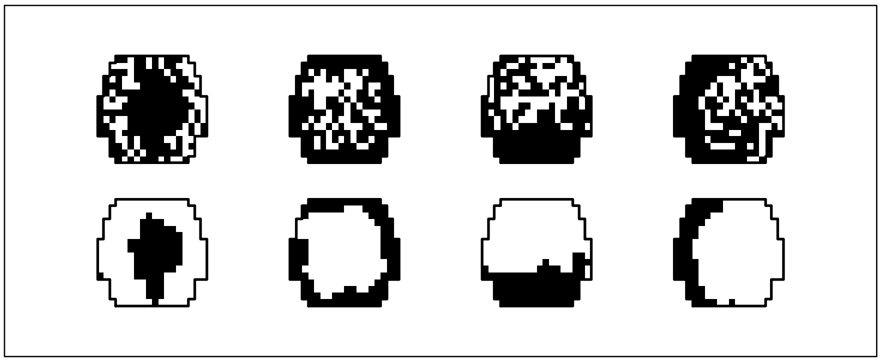

Figure 8. *Optimally chosen thresholds: 0.88, 0.74, 0.88 and 0.94, respectively, using a critical value of 0.6.*

In the next section, we describe a series of simulations designed to systematically compare these two procedures.

In Hansen et al. [1996] we describe how tests for spatial randomness based on the join-count statistics can be altered to allow for low levels of clustering among what we are referring to here as small-area defects. We model this mild defect clustering as a Markov random field, letting the engineer set a level of *acceptable* patterning. It is straightforward to apply these changes when calculating our Z-scores in the threshold profile, making our threshold selection procedure more robust. For ease of presentation, we will not consider these extensions, but will focus on our two component Bernoulli set up introduced above. The interested reader is again referred to Hansen et al. [1996] for details.

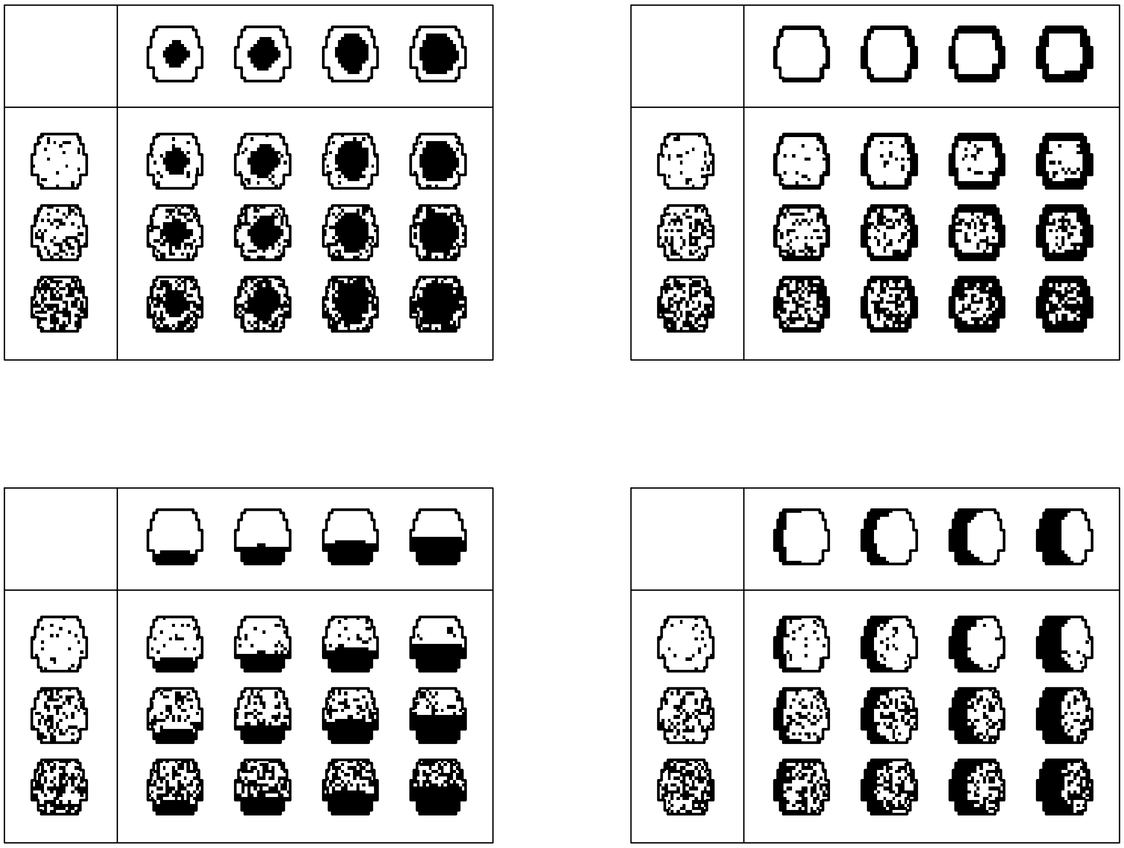

Figure 9. *Graphical description of our simulations.*

5. A simulation study

5.1 Large-area defects

Recall the framework in Figure 3 where the pattern of defective chips on a given wafer map is caused by a mixture of two spatially homogeneous Bernoulli processes. 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 3, we will consider three other cluster signatures. With varying intensities, these patterns are plotted in the top row of each table in Figure 9. By design, each set of four patterns corresponds to π_l values of 0.8, 0.7, 0.6, and 0.5, respectively. Each of these patterns is 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 handling problems, design issues or problems with wafer flatness. 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.

5.2 Small-area defects

In addition, we consider three different intensity levels for the spatially random defects. These are also shown in Figure 9. Along the leftmost row of each table we present binary maps like the first map in Figure 3 illustrating the random defect process with π_s 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 3.

5.3 The simulations

We generated 100 binary wafer maps corresponding to each entry in the tables of Figure 9. For each map, two estimates of π_l 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 9 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 9 and generated (possibly) two sets of 100 wafers per cell.

6. Simulation results

6.1 Uniformly distributed defects

Suppose that each chip on a wafer is defective with probability π_s, independent of the status of the remaining chips. Under this simple model, defects are said to be distributed uniformly across the wafer and hence π_l is taken to be 1. We begin our simulation study by examining the behavior of the estimates of π_l in this context. For each of fourteen different values of π_s shown in Figure 10, we generated 250 binary wafer maps. Boxplots of the distributions for the two estimates of π_l are given in Figure 10. 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 π_l 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.

This example is included mainly to indicate how the procedures compare under ideal conditions. As mentioned above, in Hansen et al. [1996], we develop a complete framework for performing these calculations when the spatial distribution of defective chips is described by a Markov random field. This model allows for mild spatial clustering in what we have been referring to as the small-area or random defect mechanism. From the point of view of our thresholding procedure, this implies that certain simple corrections have to be made to the null distribution of our test statistics based on an *acceptable* level of clustering. Rather than complicate issues by introducing simulation results under these models, the reader is referred to the paper Hansen et al. [1996] for a general discussion of defect clustering and process monitoring.

6.2 Area defects

We now examine the performance of the estimates of π_l when genuine area defects are present. Recall that each of the four basic patterns were run at four levels of π_l (0.8, 0.7, 0.6, 0.5) and three levels of background noise (0.9, 0.7, 0.5) as described in Figure 9 of Section 5.1. In Figures 11a-c, we present simple plots of the average computed over 100 simulations of the window estimate of π_l 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_s with the convention that the darker the color, the lower the value of π_s. In addition, within each plot, points corresponding to the same value of π_s are joined by a line. This should help the reader differentiate between the effect of varying π_l and that of varying π_s. Tables of these estimates and associated standard errors are given in the Appendix.

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

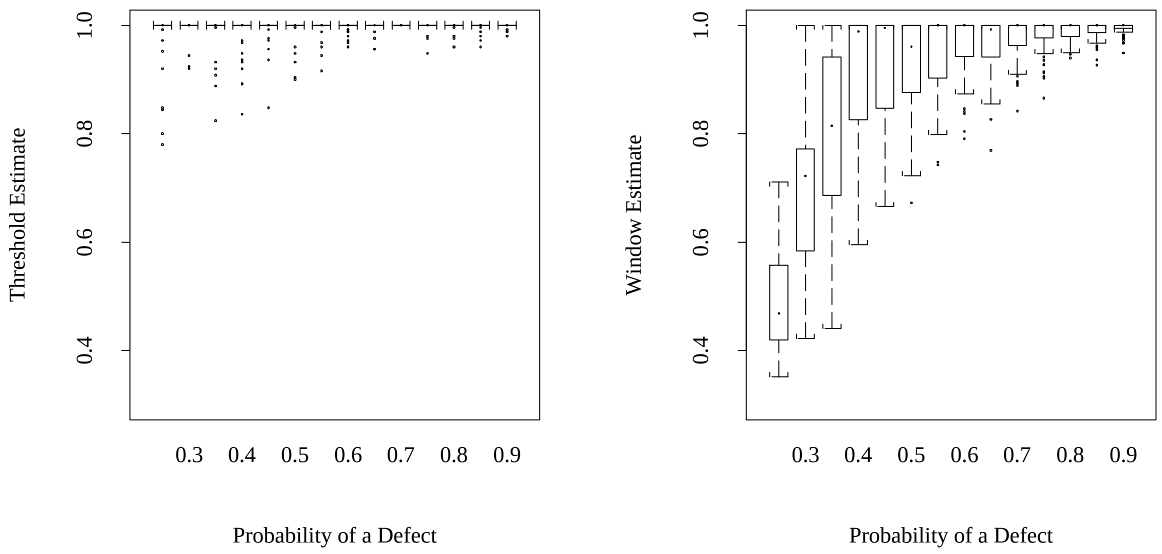

Figure 10. *The two estimates of* π_l *compared for several values of* π_s.

difficulty distinguishing between this model and a uniform defect distribution for low values of π_s. 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 9). 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 3, both techniques are biased. From the standpoint of our procedure, this bias is worse for models with extremely low values of π_s. Basically, as the yield of the random defect mechanism decreases, we incorrectly classify more and 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 11a-c. It would appear that in some cases, this estimate actually improves as

π_s 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 π_l 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_s. 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_s, 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 12 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 9. In each, the solid line describes the behavior of the threshold estimate of π_l 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 π_l for these simulations.

7. Concluding remarks

We have described a model-free approach for estimating π_s and π_l based on probe test results at the wafer map level. The approach directly exploits the positions of failing chips within a wafer to identify clusters of defective chips. An important byproduct of this analysis is a decomposition of wafer yield that attributes defective chips to large-area, clustered defects or small-area, random defects. As shown through the simulation study, the method compares favorably with the windowing method currently available in the literature and has

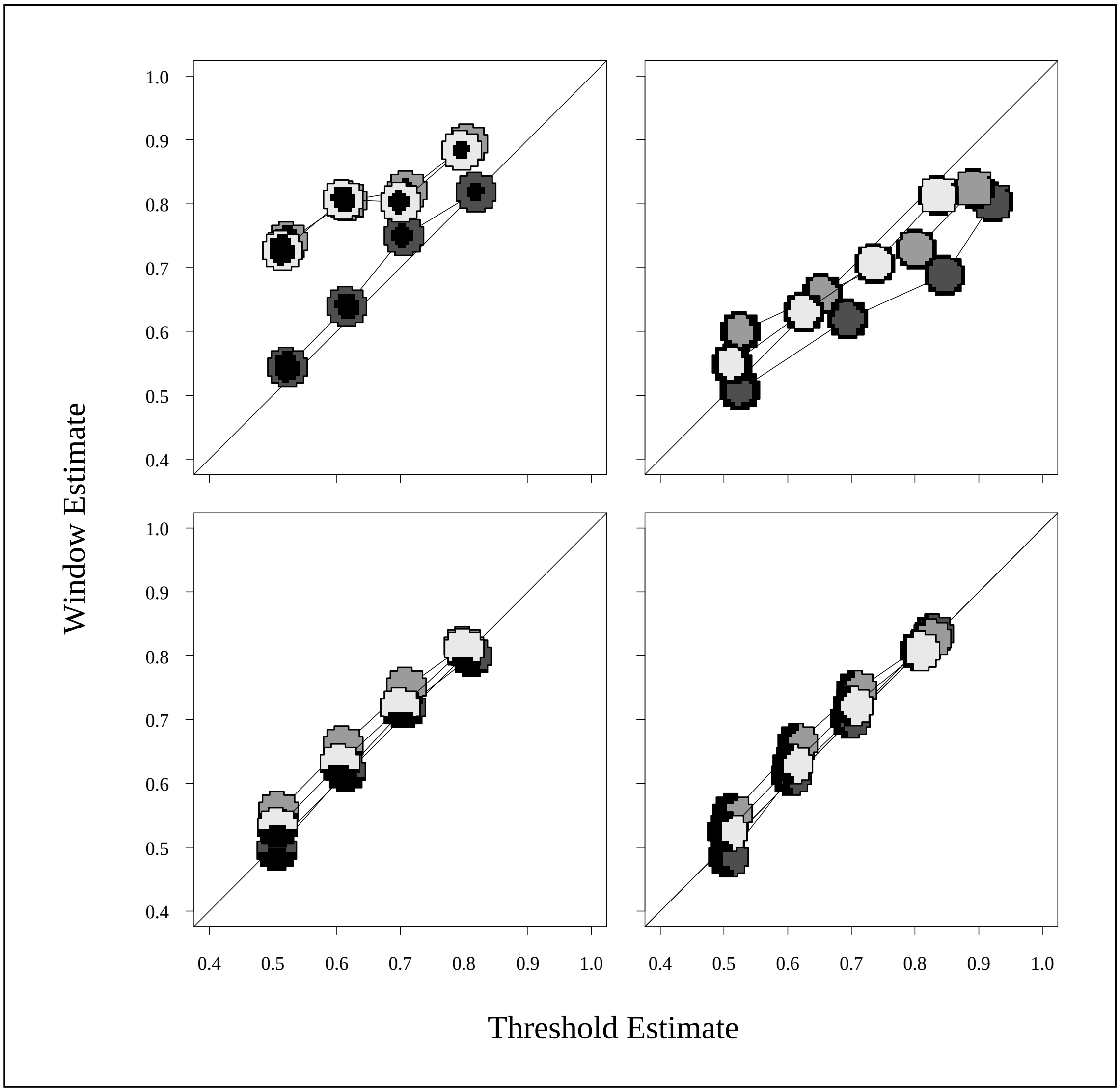

Figure 11a. *Comparison of window and threshold estimates, 100 chips per wafer.*

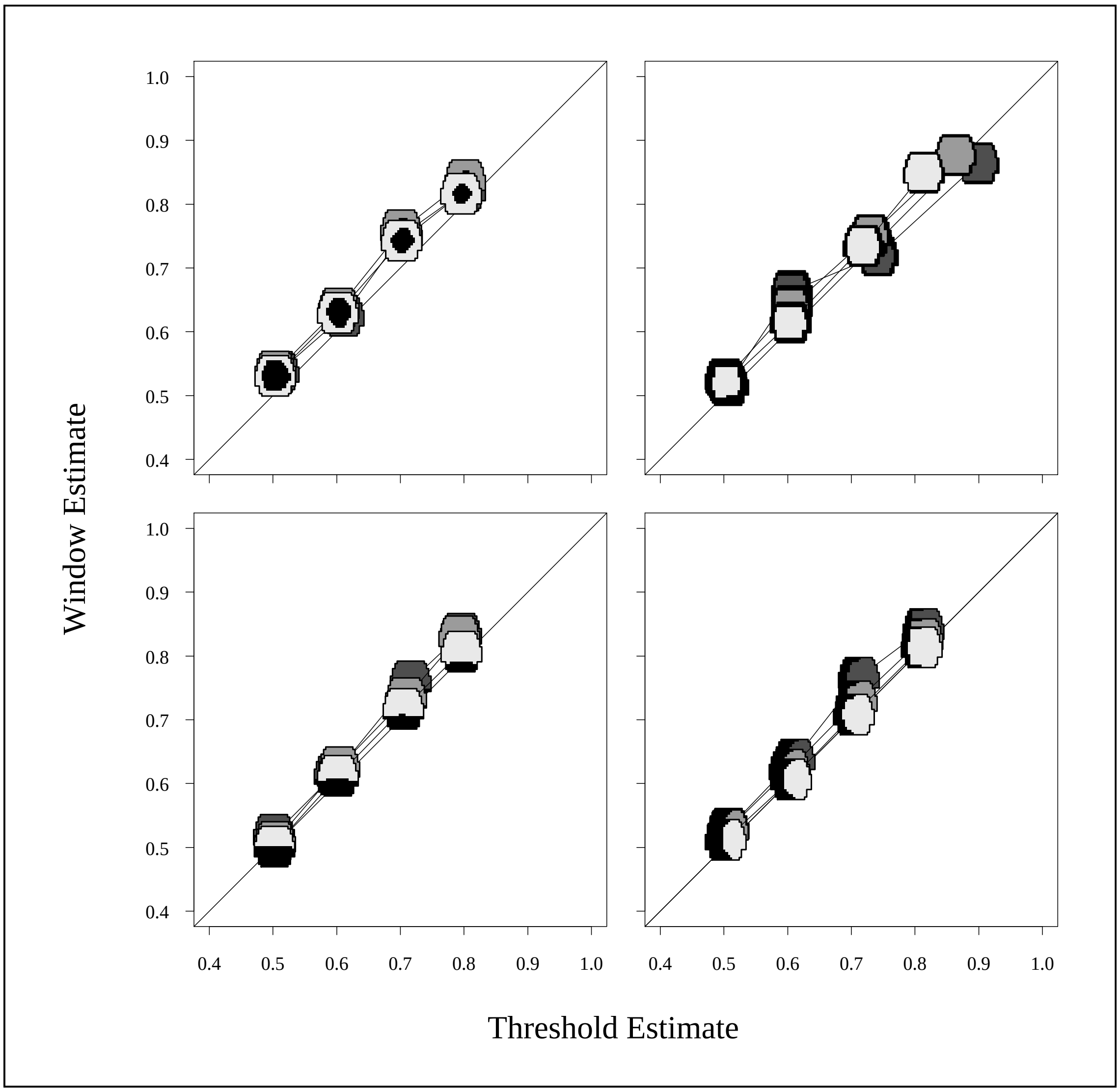

Figure 11b. *Comparison of window and threshold estimates, 250 chips per wafer.*

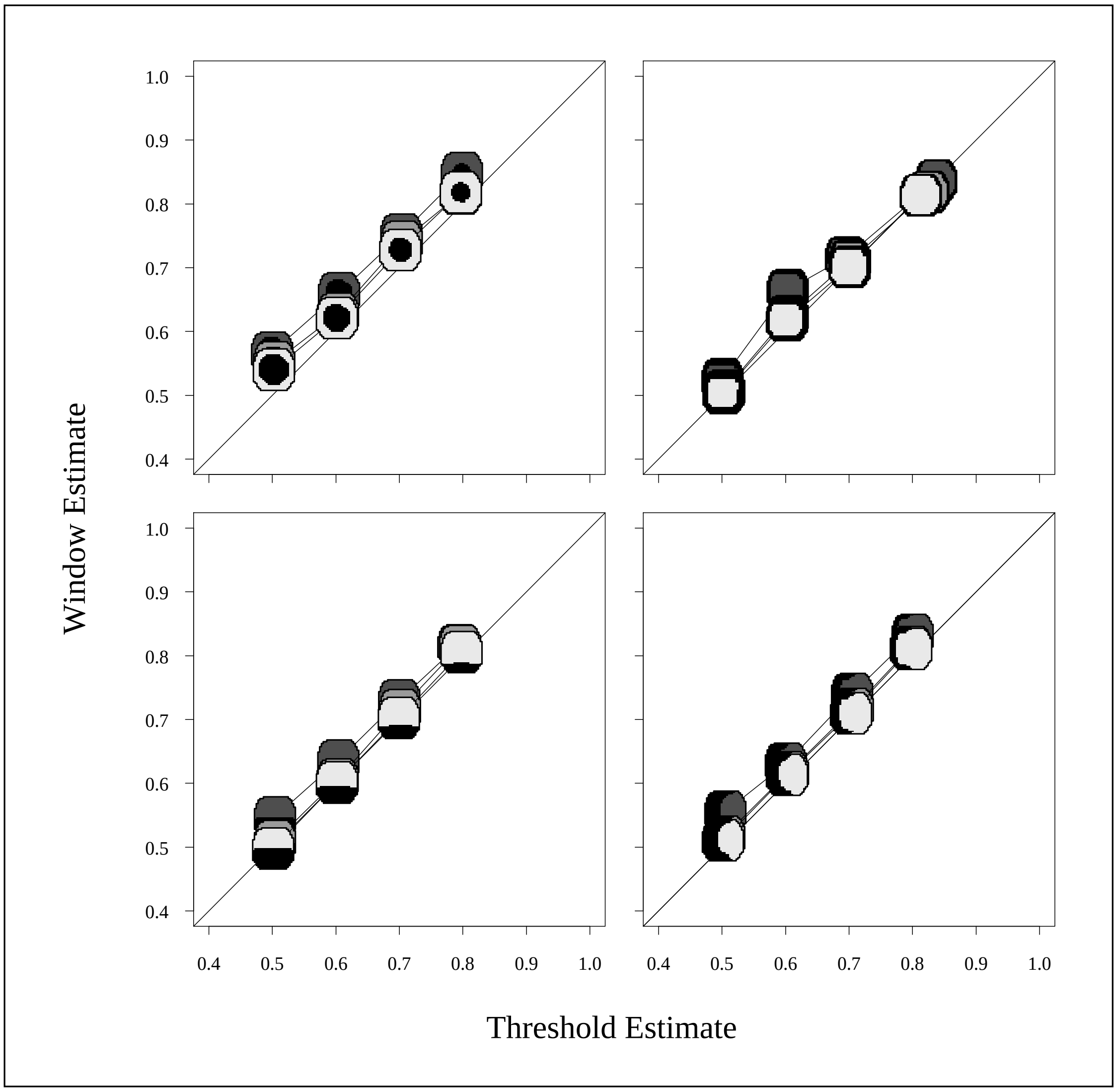

Figure 11c. *Comparison of window and threshold estimates, 500 chips per wafer.*

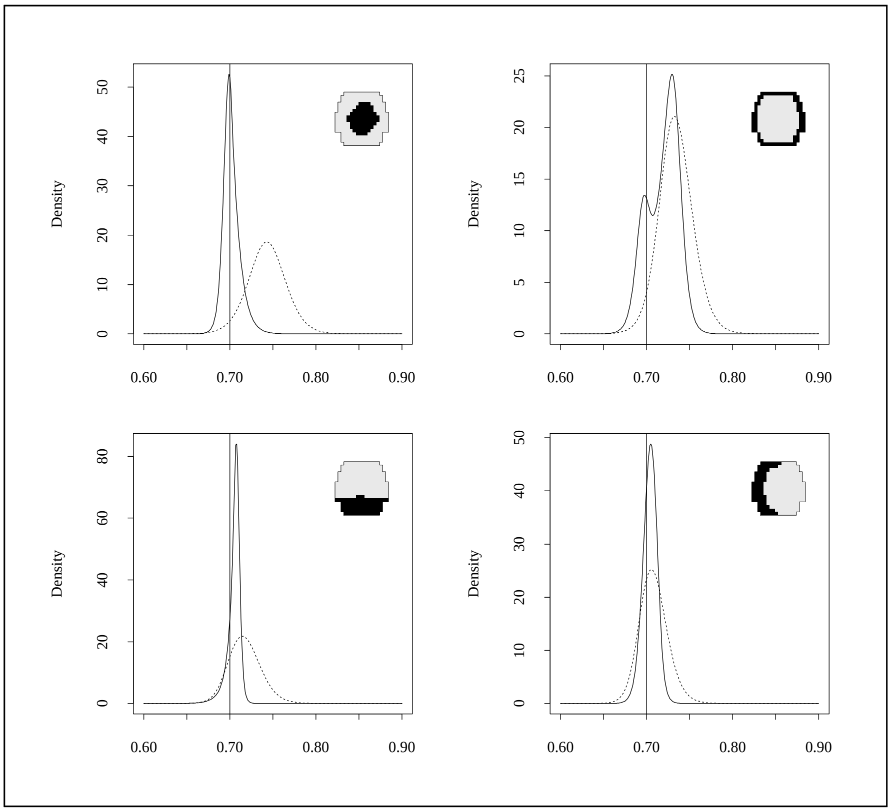

Figure 12. *Distribution of the estimates for various simulations.*

the additional advantage of producing estimates of the shape, size and location of clusters.

As noted earlier, our interest in estimating these quantities comes from a desire to better monitor the manufacturing process. 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, specifically Figures 1 and 2, will be useful in addressing this concern. More refined control charts based on the estimated values of π_s and π_l can be developed and used to track changes in defect density and clustering over time. Because these parameters depend strongly on the type of device being manufactured (its technology and code), this monitoring scheme will only be useful for products of the same 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 to estimate π_s and π_l using the methods described in this paper and model 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.

Finally, we have only considered mixtures of spatially homogeneous Bernoulli processes having two components, one associated with a low (essentially zero) probability that a given chip is fault-free, and another for which the same probability is (relatively) high. As mentioned in Section 2, our procedure can be adapted to the case in which the two probabilities are simply different, not just different with one close to zero. This can be implemented easily following the outline of the thresholding procedure given in Section 4. In this case, however, our conditioning to identify an appropriate threshold will involve both good and bad chips. For more details on the distribution of the test statistics in this type of situation, the reader is referred to Hansen et al. [1996].

8. Acknowledgements

The authors would like to thank J. R. Hoyer, Daryl Pregibon, Steve Whitlock, Bill Troutman, and Dave Anderson for their comments and suggestions. V. N. Nair's research was supported by NSF Grant DMI-9501217 and by AFOSR/ARPA MURI grant F49620-95-1-05.

9. References

[1] Cliff, A. and Ord, J. [1981] "Spatial Processes: Models & Applications," Pion, London.

[2] Cunningham, J.A. [1990] "The Use and Evaluation of Yield Models in Integrated Circuit Manufacturing," IEEE Trans. Semicond. Manufact. Vol. 3, No. 2, p. 60-71.

[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 [1996] "Identification of Spatial Patterns of Defect Clustering in Wafer Map Data," (in preparation).

[5] Hansen, M.H., V.N. Nair, D.J. Friedman [1996] "Monitoring Wafer Map Data from Integrated Circuit Fabrication Processes for Spatially Clustered Defects,"
(Submitted to *Technometrics*).

[6] Koren, I. , Koren, Z. and Stapper, C.H. [1994] "A Statistical Study of Defect Maps of Large Area VLSI ICs," IEEE Trans. on VLSI Systems, Vol. 2, No. 2, p. 251-256.

[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] Stapper, C.H. [1989] "Large-area Fault Clusters and Fault Tolerance in VLSI Circuits: A Review," IBM J. of Res. Develop., 33, 2.

[17] Taam, W. and M. Hamada [1992] "Detecting Spatial Effects from Factorial Experiments: An Application in Integrated- Circuit Manufacturing," Technometrics, Vol. 35, p. 149-160.

[18] Warner, R.M. [1981] "A note on IC-yield statistics," Solid- State Electronics, Vol. 24, No. 11, p. 1045-1047.

10. Appendix

We begin this appendix by recalling 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 3, 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

This material is discussed at length in Hansen et al. [1996] but portions are included here for completeness. 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

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

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 $(T_{\mathcal{D}}, T_{\mathcal{W}\backslash\mathcal{D}})$ 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 $T_{\mathcal{W}\backslash\mathcal{D}}$ 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

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

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 $T_{\mathcal{W}\backslash\mathcal{D}}$ can be approximated by a bivariate normal distribution with the above moments.

Tables of Results

		π_l			
		0.8	0.7	0.6	0.5
π_s	0.9	0.798	0.701	0.609	0.513
		(0.014)	(0.016)	(0.020)	(0.020)
	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)

		π_l			
		0.8	0.7	0.6	0.5
π_s	0.9	0.839	0.735	0.620	0.513
		(0.033)	(0.041)	(0.028)	(0.018)
	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)

		π_l			
		0.8	0.7	0.6	0.5
π_s	0.9	0.799	0.700	0.605	0.507
		(0.010)	(0.015)	(0.016)	(0.014)
	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)

		π_l			
		0.8	0.7	0.6	0.5
π_s	0.9	0.808	0.703	0.610	0.508
		(0.016)	(0.012)	(0.015)	(0.014)
	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 π_l and their standard errors, 100 chips per wafer. Tables are referenced to Figure 9.*

		π_l			
		0.8	0.7	0.6	0.5
π_s	0.9	0.892	0.806	0.797	0.717
		(0.041)	(0.048)	(0.053)	(0.080)
	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)

		π_l			
		0.8	0.7	0.6	0.5
π_s	0.9	0.815	0.706	0.624	0.548
		(0.031)	(0.033)	(0.028)	(0.031)
	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)

		π_l			
		0.8	0.7	0.6	0.5
π_s	0.9	0.812	0.719	0.628	0.529
		(0.028)	(0.027)	(0.028)	(0.025)
	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)

		π_l			
		0.8	0.7	0.6	0.5
π_s	0.9	0.809	0.725	0.637	0.524
		(0.028)	(0.033)	(0.035)	(0.026)
	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 π_l and their standard errors, 100 chips per wafer. Tables are referenced to Figure 9.*

		π_l			
		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)
π_s	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)

		π_l			
		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)
π_s	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)

		π_l			
		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)
π_s	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)

		π_l			
		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)
π_s	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* π_l *and their standard errors, 250 chips per wafer. Tables are referenced to Figure 9.*

		π_l			
		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)
π_s	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)

		π_l			
		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)
π_s	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)

		π_l			
		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)
π_s	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)

		π_l			
		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)
π_s	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* π_l *and their standard errors, 250 chips per wafer. Tables are referenced to Figure 9.*

		π_l			
		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)
π_s	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)

		π_l			
		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)
π_s	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)

		π_l			
		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)
π_s	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)

		π_l			
		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)
π_s	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* π_l *and their standard errors, 500 chips per wafer. Tables are referenced to Figure 9.*

		π_l			
		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)
π_s	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)

		π_l			
		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)
π_s	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)

		π_l			
		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)
π_s	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)

		π_l			
		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)
π_s	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* π_l *and their standard errors, 500 chips per wafer. Tables are referenced to Figure 9.*