# Monitoring Wafer Map Data from Integrated Circuit Fabrication Processes for Spatially Clustered Defects

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

Vijayan N. Nair
*University of Michigan*
*Ann Arbor, Michigan*

David J. Friedman
*AT&T Bell Laboratories*
*Princeton, New Jersey*

**Abstract**

Quality control in integrated circuit (IC) fabrication has traditionally been based on overall summary data such as lot or wafer yield. These measures are adequate if the defective IC's are distributed randomly both within and across wafers in a lot. In practice, however, the defects often occur in clusters or display other systematic patterns. In general, these spatially clustered defects have assignable causes that can be traced to individual machines or to a series of process steps that did not meet specified requirements. In this paper, we develop methods for routinely monitoring probe test data at the wafer map level to detect the presence of spatial clustering. The statistical properties of a family of monitoring statistics are developed under various null and alternative situations of interest, and the resulting methodology is applied to manufacturing data.

*Key Words:* Join-Count Statistic; Markov Random Field; Quality Improvement; Spatial Randomness; Statistical Process Control.

# 1 Introduction

The manufacture of integrated circuits (IC's) is a complex and costly process involving hundreds of separate steps and lasting up to twelve weeks. Several hundred IC's, or chips, are fabricated simultaneously on a *wafer*, and the wafers themselves are processed together in groups called *lots*. Upon completion of the fabrication process, each chip on every wafer is subjected to a series of functionality tests and is declared to be either good or defective. The results of these *final probe tests* can be displayed in the form of a binary wafer map. In Figure 1, we present a sample of the probe data collected from three different lots. In each case, black squares denote defective chips, while the remaining white squares represent good chips.


**Figure 1.** *Samples of wafers from three lots: a) the wafers in the first lot had a high yield and did not suffer from excessive defect clustering (233 chips/wafer); many of the wafers in the second lot had central regions of zero yield (300 chips/wafer); and finally, c) over half of the wafers in the third lot had a thin, crescent shaped defect pattern (210 chips/wafer).*

Analysis of probe test data for quality control and process monitoring purposes has been traditionally restricted to overall lot-level summary measures such as *yield* (the number of good chips in a lot) and the *good-to-functional* ratio (the number of good chips in a lot divided by the number of chips that work but do not meet specification limits). Such overall measures are adequate only if defects are distributed randomly both within and across wafers in the lot. While the sample of wafers in the top row of Figure 1 seem to exhibit this property, we have found that it is common for defective chips to occur in clusters or to display other systematic patterns, as illustrated in the remaining two rows of Figure 1.

Spatial patterns such as these contain useful information about potential manufacturing problems that is missed by overall summary measures. For example, an uneven temperature distribution during the rapid thermal annealing (RTA) process can produce a ring of dead chips around the edge of the wafer; excess vibration in a given machine might liberate enough particles to cause all the chips in some contiguous region of a wafer to fail; and it is common for a malfunctioning stepper to generate a checkerboard pattern of defective chips. In general, clusters of defects can be classified as either particle or process related, with particle related clusters being assignable to individual machines and process related clusters being attributable to one or more process steps not meeting specification requirements.

On the other hand, spatially random defects like those in the first row of Figure 1 are often the product of the cleanroom environment. Random defect densities tend to rise and fall with the overall cleanliness of the cleanroom, 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, materials-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. These problems, however, will be picked up by the existing yield summaries at either the lot or wafer level.

In general, the pattern of defective chips on a given wafer map is caused by a superposition of two destructive mechanisms. Acting independently, the first generates defects randomly on the wafer, while the second produces clusters of defects. This model is depicted graphically in Figure 2. In terms of traditional statistical process control terminology, the first mechanism can be viewed as a random or common cause while the second is a special or assignable cause. The goal of this paper is to develop methods for routinely monitoring wafer map data to detect the presence of significant spatially clustered defects; that is, to detect the presence of the second defect mechanism in Figure 2. In practice, however, almost all wafer map data exhibits some form of mild spatial clustering, and hence the notion of a completely random defect mechanism is an idealization. Therefore, we must also consider the more general problem of detecting significant clusters of defects in the presence of mild spatial inhomogeneity. Toward this end, we will study the properties of our monitoring procedure under null hypotheses based on Markov random field models.

The procedures we propose are intended to augment current monitoring efforts that focus on overall summary measures like yield, calculated at either the wafer or the lot level. Once a set of wafers has been tagged, we can subject them to further statistical and root-cause analyses to identify the nature of the defects and ultimately use this information to improve the process. Friedman et al. (1996) describe a particular statistical strategy for analyzing tagged lots that maintains the two-dimensional character of data collected at final probe and systematically exploits the information contained in the spatial distribution of defects across wafers.

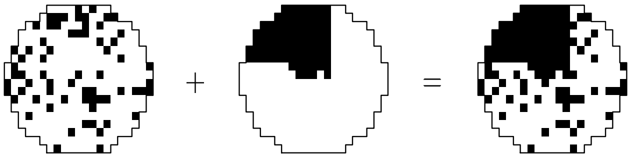


**Figure 2.** *A graphical model of the two destructive forces working independently to create the observed wafer map. This layout has 298 chips/wafer.*

The spatial clustering phenomenon has been discussed extensively in the IC fabrication literature (see, for example, Mallory et al., 1983). However, very little work has been done to systematically exploit this information for yield and process improvement. Most of the existing literature is focused on examining the effect of clustering on yield prediction (Cunningham, 1990). Albin and Friedman (1989), for example, proposed control charts for monitoring yield in the presence of spatial clustering. Their work is based on fitting "over-dispersion" models such as the negative-binomial or Neyman-Type A distributions to account for the excess variation and use these to develop appropriate control limits. Our approach is quite different in that sources of over-dispersion are not merely accounted for, but rather are viewed as the starting point for effective yield improvement efforts. While yield models in the IC literature typically only consider data at a lot level, whenever a wafer-level analysis has been proposed, it has ignored any spatial information (see Mallory et. al., 1983). Taam and Hamada (1993), on the other hand, considered a wafer-level analysis for data collected from single-lot, designed experiments in order to assess the extent of clustering as a function of the experimental factors under investigation. Our monitoring procedures are based on statistics that are related to those used by Taam and Hamada (see also Cliff and Ord, 1981). However, in contrast to this single-lot, wafer-level analyses, we are concerned with techniques for routinely monitoring post-manufacturing wafer map data to detect the presence of spatial clustering.

The rest of the paper is organized as follows. In order to make this discussion accessible to readers who may not be interested in the technical development of the methodology, we present a summary of our approach and illustrate its application in Section 2. In Section 3 we motivate our monitoring procedure by introducing a rich family of tests based on local estimates of defect probabilities. The monitoring statistic we propose is a member of this family. The performance of the monitoring statistic is studied under a variety of spatial defect distributions in Section 4 and its power behavior is examined in Section 5.

## 2 The monitoring procedure and applications

Since yield is the primary metric of interest in IC fabrication, it is usually monitored routinely. Unfortunately, a typical manufacturing facility produces so many wafers per week that engineers are not able to visually inspect each lot to assess trends in the type and extent of spatial clustering. Thus, statistical procedures are required to routinely monitor wafer map data and screen lots based on the presence of significant clustering. For example, there is no reason for engineers to visually examine the wafer map data in the first lot of Figure 1, and hence such lots should be screened out by the monitoring procedure. On the other hand, even if the overall yields of the second and third lots in this figure are within acceptable limits, these maps contain useful information about the manufacturing process. An examination of the clustering present in these lots and a root-cause analysis could eventually lead to process improvements, and hence the monitoring procedure should tag such lots for subsequent analysis by engineers.

In Section 3, we cast these requirements as a formal hypothesis testing problem at the individual wafer map level and develop appropriate test statistics. We will discuss later in this section operational guidelines for applying the procedure to monitor data summarized at the lot level. The particular wafer-level test statistic that we have found useful is

$$T = T_{\mathcal{N}_1} + T_{\mathcal{N}_0} ,$$

where $T_{\mathcal{N}_1}$ and $T_{\mathcal{N}_0}$ are defined in equations (3.3a) and (3.3b). This test statistic is a sum of two weighted "join-count" statistics; one counts the number of "defective" neighbors of defective chips and the other counts the number of "non-defective" neighbors of non-defective chips. For a fixed number of defective chips in the wafer, the value of the test statistic will be large if either the defective or the non-defective chips tend to be spatially clustered.

The properties of this test statistic are systematically developed in Sections 4 and 5. In particular, its first two moments under both the null hypothesis of complete spatial randomness and some clustered alternative hypotheses are derived. It is also shown that the distribution of this test statistic can be reasonably approximated in situations of interest by a normal distribution with the corresponding moments. Thus, we can use large sample normal approximations to develop appropriate critical values and study its power behavior. As noted earlier, the hypothesis of complete spatial randomness is a mathematical idealization, and in practice, one typically observes some mild spatial inhomogeneity in the underlying defect intensity. For example, the defect probabilities tend to increase near the edge of the wafer map. Thus, a more realistic testing scenario is to view some form of mild spatial inhomogeneity as the null hypothesis of interest. In Section 5, we demonstrate that our proposed test statistic has good power in detecting significant clusters of defects under these more realistic conditions.

We now illustrate the application of our procedure to some real data sets. Figure 3a shows the (binary) probe test data for a lot of 20 wafers. It is clear from a visual examination of these data that the defective chips in this lot are not spatially clustered. In Figure 3b we present one possible graphical implementation of our monitoring technique applied to these 20 wafers. The plot itself is just a $p$-chart for monitoring yield, where yield is plotted against wafer number and we have shaded the region between the upper and lower control limits. Information about defect clustering is encoded in the plotting character: the symbol '*' is used to denote wafers that exhibit significant spatial clustering (corresponding to a given statistical significance level, in this case 0.05), while the character 'o' is used to mark those wafers that are not significantly clustered. As expected, we find that there is very little spatial clustering in this group of 20 wafers. Our interactive implementation of this tool allows the user to point-and-click on any particular symbol and display the corresponding wafer map at the lower left hand corner of the plot. In Figure 3b, for example, we have selected wafer number 12. This type of interaction allows the user to visually calibrate the results of the test as well as to assess the extent and nature of spatial clustering present in

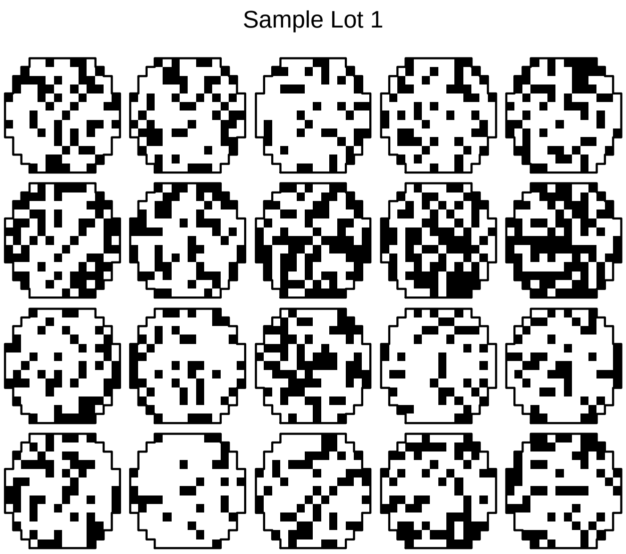


**Figure 3a.** *A lot of 20 wafers that do not seem to exhibit much spatial clustering (160 chips/wafer).*

each wafer. (When used in the form of a report, we replace the symbols '*' and 'o' with the actual wafer maps, and encode clustering information by coloring the wafer borders. In addition, when attempting to relate spatial patterning with processing information, we often order the wafers according to their sequence through a given machine or fabrication step.) Figures 4a and 4b show the application of this tool to a lot of 20 wafers that exhibit considerable defect clustering. In this case, many of the wafers are tagged with '*'s, indicating the presence of significant, clustered defects.

This last example highlights the interplay between our monitoring procedure and the more traditional $p$-chart constructed for wafer yield: wafer number 14 has a very low yield, but does not exhibit much spatial patterning, while the small number of defective chips on wafer 6 are tightly clustered. In addition, this particular display has certain advantages simply because it is an enhanced version of the $p$-chart, a tool already used extensively by engineers to monitor yield, and the additional information on spatial clustering is presented non-intrusively. The interactive feature allows the user to examine the data in more detail if she so chooses. However, this type of tool requires the user to visually examine the figure for each lot and decide whether to tag a lot for further analysis. It is possible to automate

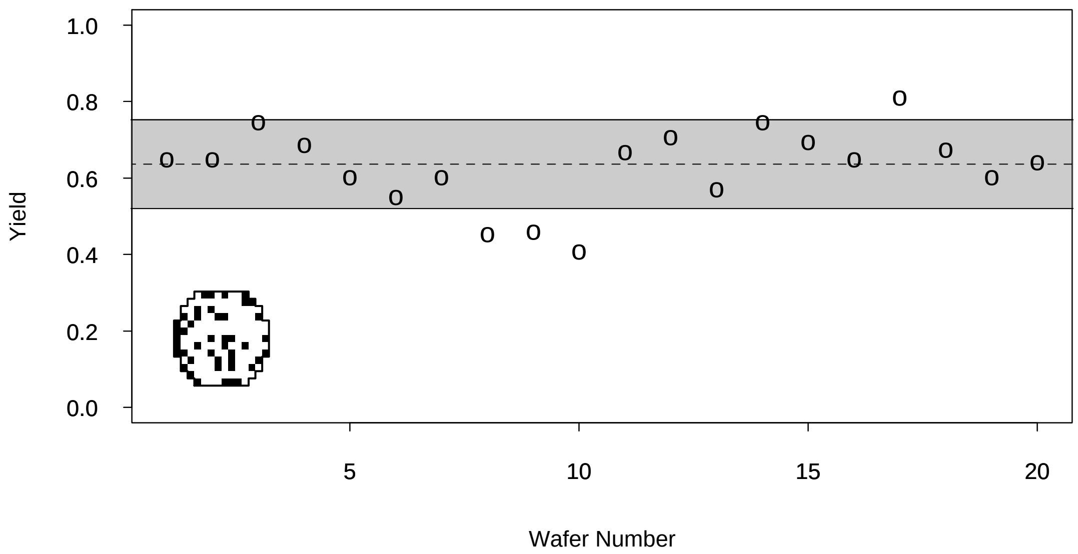


**Figure 3b.** *Applying the proposed monitoring procedure to the wafers in Figure 3a. The shaded region is bounded by the upper and lower control limits, and the centerline is dashed. Points marked with the symbol 'o' represent wafers without significant spatial clustering, while those marked with a '*' identify wafers that have been tagged as containing clusters of defects.*

this procedure by developing operational guidelines for making this decision at the lot level. One such rule might tag a lot if more than a certain proportion, say 20%, of the wafers are spatially clustered at a given statistical significance level. The operating characteristics of this rule are studied in Section 5.

We can also implement our monitoring procedure using more traditional control charting techniques, either at the wafer- or lot-level. A simple chart of either the value of our test statistic or its Z-value against wafer number is just one example. Again, a rule such as the one mentioned in the previous paragraph can be used to automate decisions at the lot level. In addition, a lot-level chart can be obtained by plotting the proportion of wafers in the lot that are significantly clustered (at any given level of significance) against lot number. We can then use a centerline and control limits to monitor the proportion of significantly clustered wafers. While there are clearly a number of possible implementations, we prefer

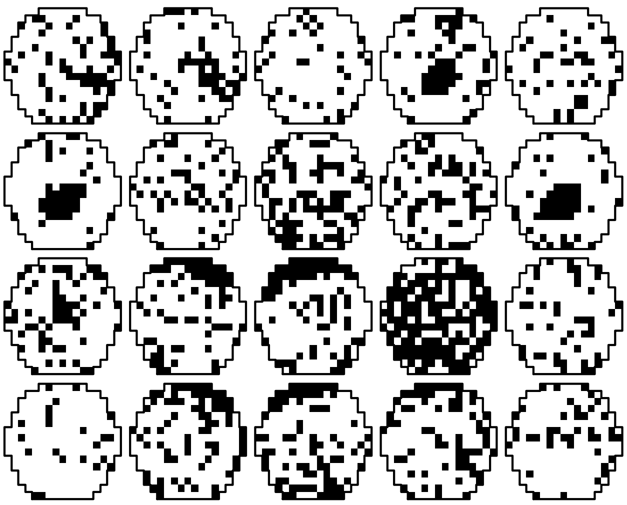


**Figure 4a.** *A lot of 20 wafers with considerable spatial clustering (220 chips/wafer).*

to let the readers customize such tools to fit their own needs. In our applications, we have found the visual displays in Figures 3b and 4b to be the most informative. However, we also recognize that other users may prefer to adapt these tools to their environments.

## 3 Which statistic should we monitor?

In Figure 2, we presented a graphical representation of a simple model for the spatial distribution of defective chips across a wafer. To make this intuitive formulation mathematically precise, we need to introduce some notation. In what follows, we use a script letter when referring to a set and let the corresponding upper case letter denote its cardinality. For example, $\mathcal{W}$ represents the collection of chip locations on a given wafer, while $W$ denotes the total number of chips on the wafer. Given a set of final probe data, we construct a binary wafer map by assigning to each chip position the value "0" if the associated chip is good and "1" if it is defective. Following this convention, let $\mathcal{N}_1 \subset \mathcal{W}$ denote the positions of the defective chips on a given wafer and set $\mathcal{N}_0 = \mathcal{W} \setminus \mathcal{N}_1$, the locations of the good chips. Hence, there are $N_1$ defective and $N_0$ good chips on a given wafer map, and $W = N_1 + N_0$. Next, for $i \in \mathcal{W}$, let $I_{\mathcal{N}_1}(i)$ be the indicator function of the event that the chip at position $i$

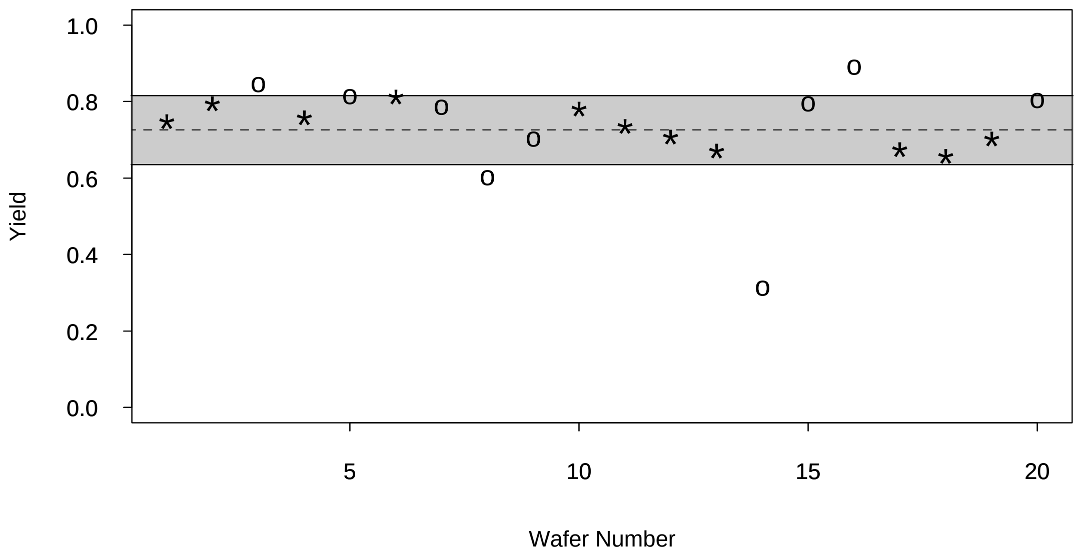


**Figure 4b.** *Applying the proposed monitoring procedure to the wafers in Figure 4a. The shaded region is bounded by the upper and lower control limits, and the centerline is dashed. Points marked with the symbol 'o' represent wafers without significant spatial clustering, while those marked with a '*' identify wafers that have been tagged as containing clusters of defects.*

is defective, and set $I_{\mathcal{N}_0}(i) = 1 - I_{\mathcal{N}_1}(j)$. Finally, we let $p_i = EI_{\mathcal{N}_1}(i)$ denote the probability that the chip at location $i$ is defective, and set $q_i = 1 - p_i$, the probability that the same chip is good.

In the absence of clustering, we assume that defective chips are distributed randomly across the wafer. Under this hypothesis of *complete spatial randomness* (CSR), the variables $I_{\mathcal{N}_1}(i)$ are independent and identically distributed as Bernoulli random variables with constant probability $p_i = p$, $i \in \mathcal{W}$; that is, we have a spatially homogeneous Bernoulli process over the chip locations in $\mathcal{W}$. In what follows, we will use the terms "spatially homogeneous" and "CSR" interchangeably to refer to an absence of clustering. The four wafer maps presented in each row of Figure 5 were simulated under the null hypothesis of CSR. As is evident from this figure, when the overall yield is low, even spatially homogeneous wafer maps can appear to have clusters of defects. These examples underscore the difficulty in detecting the presence of spatial patterns both by visual examination and by formal statistical tests. In general, when the yield is close to zero or one, the null and alternative hypotheses are not clearly distinguishable. As a practical matter, however, this is not a serious issue since wafers are routinely monitored for variations in yield, and hence detecting spatially clustered defects is only of interest for wafer maps with yield within certain limits.

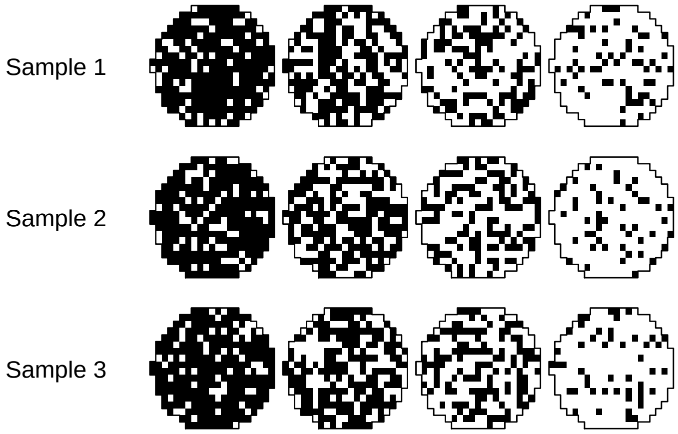


**Figure 5.** *Samples of wafers simulated under the hypothesis of CSR. The wafers in each column have yields 20%, 40%, 60% and 80%, respectively (300 chips/wafer).*

Under the null hypothesis of complete spatial randomness, each chip $i \in \mathcal{W}$ has the same probability $p$ of being defective. We can estimate $p$ either globally by $\hat{p} = N_1 \,/\, W$, or locally by

$$\hat{p}_i = \sum_{j \in \mathcal{W}} w_i(j) I_{\mathcal{N}_1}(j)\,,$$

where for each $i \in \mathcal{W}$, $\{w_i(j)\}$ denotes a set of non-negative weights normalized so that $\sum_{j \in \mathcal{W}} w_i(j) = 1$. While the form of the weights can vary from chip to chip, we will typically take them to be constant in a fixed neighborhood around the chip of interest and to be zero elsewhere. For chips near the boundary of the wafer, the neighborhood has to be modified and the weights renormalized in the obvious way to include only chip sites in $\mathcal{W}$.

When clustering is present, we expect to see considerable variation in the local estimates $\hat{p}_i$ of $p$. This observation leads us to the following class of statistics for testing the null hypothesis of spatial randomness:

$$T \;=\; \sum_{i\in\mathcal{N}_1} \phi_{\mathcal{N}_1}\,(\hat{p}_i,\ \hat{p}) \;+\; \sum_{i\in\mathcal{N}_0} \phi_{\mathcal{N}_0}(\hat{q}_i,\ \hat{q})\,, \tag{3.1}$$

where $\hat{q} = 1 - \hat{p}$ and $\hat{q}_i = 1 - \hat{p}_i$, $i \in \mathcal{W}$. Many obvious candidates for test statistics can be written in this form. For example, if we set $\phi_{\mathcal{N}_1}(x, y) = \phi_{\mathcal{N}_0}(x, y) = |x - y|^\nu$, then (3.1) becomes

$$T \;=\; \sum_{i\in\mathcal{W}} |\hat{p}_i - \hat{p}|^\nu,$$

which is equivalent to the $L_\nu$-distance between the local and global estimates of the defect probabilities. A large value of the test statistic indicates that the local estimates are quite different from the overall wafer-level estimate, suggesting that the null hypothesis of spatial homogeneity does not hold. As a second example, consider $\phi_{\mathcal{N}_1}(x, y) = \phi_{\mathcal{N}_0}(x, y) = \log(x/y)$. In this case, we find that

$$T \;=\; \sum_{i\in\mathcal{W}} I_{\mathcal{N}_1}(i)\ \log(\hat{p}_i/\hat{p}) + \sum_{i\in\mathcal{W}} I_{\mathcal{N}_0}(i)\ \log(\hat{q}_i/\hat{q}).$$

This statistic is structurally the same as the likelihood ratio test of a simple null hypothesis of spatial randomness against a completely specified, spatially nonhomogeneous Bernoulli process. Under this alternative, the variables $I_{\mathcal{N}_1}(i)$ are still independent, but the probabilities $p_i = EI_{\mathcal{N}_1}(i)$ are allowed to vary from chip to chip. In our case, however, the defect probabilities are unknown and have been replaced by their estimators.

Finally, if we let $\phi_{\mathcal{N}_1}(x, y) = W^{-1}(x - y)$ with $\phi_{\mathcal{N}_0}(x, y) = 0$, we get

$$W^{-1}\sum_{i\in\mathcal{W}} I_{\mathcal{N}_1}(i)\ \hat{p}_i - \hat{p}^2 \;=\; W^{-1}\sum_{i\in\mathcal{W}}\sum_{j\in\mathcal{W}} w_i(j) I_{\mathcal{N}_1}(i) I_{\mathcal{N}_1}(j) - \hat{p}^2. \tag{3.2}$$

An analogous statistic can be obtained for the "good" chips by defining the function $\phi_{\mathcal{N}_0}(x, y) = W^{-1}(y - x)$ with $\phi_{\mathcal{N}_1}(x, y) = 0$. The resulting statistic is obtained by replacing $I_{\mathcal{N}_1}(i)$ and $\hat{p}$ in (3.2) with $I_{\mathcal{N}_0}(i)$ and $\hat{q}$ respectively. These statistics are generalized versions of Moran's (1948) test for spatial dependence and have been discussed by Cliff and Ord (1981) (see also Taam and Hamada, 1993).

If the weights are a nonzero constant only for the immediate neighbors of chip $i$, the quantity in (3.2) can be interpreted as an estimate of the *first-order* spatial covariance $EI_{\mathcal{N}_1}(i)I_{\mathcal{N}_1}(j) - EI_{\mathcal{N}_1}(i)EI_{\mathcal{N}_1}(j)$, where chip $j$ is an immediate, or first-order, neighbor of chip $i$. (This is meaningful only for spatially stationary processes.) Under the null hypothesis of CSR, the expected value of this statistic is close to zero. If the defects are spatially dependent, as is the case when defective chips occur in clusters, the statistic will take on a large, positive value. An unduly large, negative value, on the other hand, might suggest a systematic pattern of defects.

It is clear from the above discussion that one might consider other tests corresponding to higher order spatial covariances by a suitable choice of the weights $\{w_i(\cdot)\}$. Venturing outside the class given in (3.1), one can also entertain higher order moments and define tests that involve products of more than two indicator functions. Interestingly, the *windowing technique*, a regression procedure known in the IC manufacturing literature for assessing the extent of spatial clustering, is based on statistics involving these higher order moments. Simulation studies indicate, however, that subject to the constraints imposed by our particular application, this type of test does not have more power to detect clustered alternatives than the members of the class introduced in (3.1) (see Friedman, et al. 1995).

For the remainder of this paper, we limit our focus to the two test statistics of the form (3.2). More specifically, we will consider monitoring linear combinations of the following quantities:

$$T_{\mathcal{N}_1} = W^{-1} \sum_{i \in \mathcal{W}} \sum_{j \in \mathcal{W}} w_i(j) I_{\mathcal{N}_1}(i) I_{\mathcal{N}_1}(j) \tag{3.3a}$$

and

$$T_{\mathcal{N}_0} = W^{-1} \sum_{i \in \mathcal{W}} \sum_{j \in \mathcal{W}} w_i(j) I_{\mathcal{N}_0}(i) I_{\mathcal{N}_0}(j)\,. \tag{3.3b}$$

These statistics are equivalent to (3.2) when we condition on the observed value of $\hat{p} = N_1 \,/\, W$, the proportion of defective chips on the observed wafer map (see the discussion below). One motivation for this choice is that the relatively simple form of these statistics allows us to compute various moments exactly under a variety of interesting, practically important null and alternative hypotheses.

# 4 Spatial defect distributions and the pair $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$

## 4.1 The null hypothesis of complete spatial randomness

Because wafers are currently screened on the basis of yield, we will test for spatial homogeneity conditionally on the observed value of $\hat{p} = N_1 \,/\, W$, the proportion of defective chips on the observed wafer map. In terms of our spatial model, complete spatial randomness then means that the $N_1$ defective chip locations are selected at random without replacement from $\mathcal{W}$. The distribution of $T_{\mathcal{N}_1}$ and $T_{\mathcal{N}_0}$ under this hypothesis has been considered in a variety of contexts, and a comprehensive treatment can be found in Chapter 2 of Cliff and Ord (1981). To get the exact distribution of these statistics, we have to resort to the permutation distribution based on all possible allocations of $N_1$ defects among the $W$ different chips on the wafer $\mathcal{W}$. While it is not difficult to generate wafer maps that follow this distribution, it would be computationally impractical to depend on these samples for approximate critical values. By appealing to the theory of U-statistics, however, it can be shown that once suitably normalized, our pair of test statistics $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$ has a bivariate normal limiting distribution as $W$ (and hence $N_1$) tends to infinity (see Lee, 1990).

Furthermore, given $\hat{p} = N_1 \,/\, W$, it is not difficult to derive the exact first and second moments of these statistics under CSR. For example, a straightforward calculation yields

$$E\left(T_{\mathcal{N}_1} \mid \hat{p} = \frac{N_1}{W}\right) = \frac{N_1\,[N_1 - 1]}{W\,[W-1]}$$

and

$$E\left(T_{\mathcal{N}_0} \mid \hat{p} = \frac{N_1}{W}\right) = \frac{[W - N_1]\,[W - N_1 - 1]}{W\,[W-1]}\,,$$

assuming that $w_j(j)$ is zero for all $j \in \mathcal{W}$ (see also Cliff and Ord, 1981). 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 $I_{\mathcal{N}_1}(j)^2 = I_{\mathcal{N}_1}(j)$. Next, tedious computations that are closer in spirit to accounting than algebra lead us to the second moments of the pair $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$:

$$\begin{aligned} E\left(T^2_{\mathcal{N}_1} \mid \hat{p} = \frac{N_1}{W}\right) &= \alpha_1 \frac{N_1}{W} + \alpha_2 \frac{N_1\,[N_1-1]}{W\,[W-1]} + \alpha_3 \frac{N_1\,[N_1-1]\,[N_1-2]}{W\,[W-1]\,[W-2]} \\ &\quad + \alpha_4 \frac{N_1\,[N_1-1]\,[N_1-2]\,[N_1-3]}{W\,[W-1]\,[W-2]\,[W-3]}\,, \end{aligned}$$

where $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$, and $\alpha_1$ is zero if $w_j(j)$ is zero for all $j \in \mathcal{W}$. Not surprisingly, these expressions depend on the arrangement of chips within the wafer $\mathcal{W}$ and the weights $\{w_i(\cdot)\}$, as well as $N_1$ and $W$. Roughly speaking, these quantities are obtained by considering the proportion of terms in the expansion of the second moment that involve two, three and four unique indicators, respectively, and hence must be computed separately for each different chip layout $\mathcal{W}$. Finally, if $w_j(j)$ is zero for all $j \in \mathcal{W}$, then the covariance between $T_{\mathcal{N}_1}$ and $T_{\mathcal{N}_0}$ can be found using

$$E\left(T_{\mathcal{N}_1} T_{\mathcal{N}_0} \mid \hat{p} = \frac{N_1}{W}\right) = \alpha_4 \frac{N_1 \, [N_1 - 1] \, [W - N_1] \, [W - N_1 - 1]}{W \, [W - 1] \, [W - 2] \, [W - 3]},$$

where the constant $\alpha_4$ also appears in the equation for the second moments of $T_{\mathcal{N}_1}$ and $T_{\mathcal{N}_0}$. Again, if for some $j$, $w_j(j)$ is not zero, then lower-order products would be included in this expression.

As mentioned in the previous section, we would like to monitor the linear combination of the statistics $T_{\mathcal{N}_1}$ and $T_{\mathcal{N}_0}$ that has the greatest power to detect important departures from CSR. To make this determination, we need to study the joint distribution of these quantities under a variety of spatial models. As a first step, we can combine the exact (conditional) moments with the asymptotic normality result mentioned earlier to construct (approximate) probability contours for the pair $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$. In Figure 6 we present the mean of $T_{\mathcal{N}_1}$ plotted against the mean of $T_{\mathcal{N}_0}$ as $\hat{p} = N_1 \,/\, W$ varies from zero to one. The ellipses correspond to approximate contours based on the variance-covariance matrix calculated using the chip layout $\mathcal{W}$ associated with the wafer map featured in Figure 5, and setting $\hat{p}$ equal to 0.85, 0.70 and 0.55, respectively. Probability contours for the complementary values of $\hat{p}$, 0.15, 0.30 and 0.45, can be found by reflecting these ellipses through the line $y = x$. The slopes mentioned in this plot refer to the principal axes of the three ellipses. Observe that for a broad range of values of $\hat{p}$, these sets are all oriented similarly. In Proposition A.2 of the appendix we derive the exact expression for the slope of the principal axis of the approximate probability contour for the pair $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$ as a function of $\hat{p}$, and show that for large values of $W$, this quantity can be approximated by $(\sqrt{4 + \gamma^2} + \gamma)/2$, where

$$\gamma \approx C \left( \frac{\hat{q}}{\hat{p}} - \frac{\hat{p}}{\hat{q}} \right)$$

and the constant $C$ depends only on the configuration of the chips $\mathcal{W}$ and our choice of weights $\{w_i(\cdot)\}$. The beauty of this expression is that one can separate yield information from the specifics of the wafer map and neighborhood structure. For example, using the value of $C$ calculated from the chip layout of the wafer map in Figure 5, as $\hat{p}$ varies from 0.15 to 0.85, $\gamma$ spans the interval $[-0.2, 0.2]$, and hence the corresponding slopes vary from 0.81 to 1.23 as indicated in the figure. From this analysis we find that the slope of the principal axis is close to one for a large range of yield values.

This suggests that the quantity $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ given in equation (3.3) is (approximately) the first principal component of the joint distribution of the pair $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$. In the following sections, we will demonstrate that this statistic also has essentially maximal power to detect a variety of important spatial defect distributions among all possible linear combinations of $T_{\mathcal{N}_1}$ and $T_{\mathcal{N}_0}$. Strictly speaking, these statements are only valid for values of $\hat{p} = N_1 \,/\, W$ that are not too close to zero or one. For such extreme cases, it is possible to refine our arguments and let the linear combination vary with $\hat{p}$. As a practical matter, however, wafers with extremely low yield will be caught by existing monitoring procedures, while clustering in extremely high yielding wafers is of secondary importance given our current manufacturing conditions.

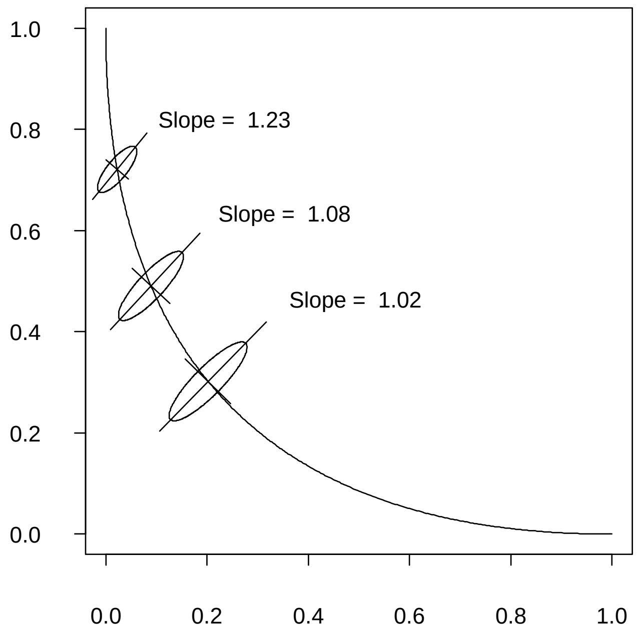


**Figure 6.** *Plot of* $E\,(T_{\mathcal{N}_1}|\hat{p})$ *versus* $E\,(T_{\mathcal{N}_0}|\hat{p})$ *for* $0 \le \hat{p} \le 1$. *Ellipses denote approximate, bivariate probability contours for the pair* $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$.

### 4.2 Distribution under clustered alternatives

In this section, we study the distribution of the pair $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$ under the full model illustrated in Figure 2. To be more precise, we will consider the behavior of these statistics conditionally on the value of $I_{\mathcal{N}_1}(i)$ for a particular region of the wafer. Let $\mathcal{C}_0, \mathcal{C}_1 \subset \mathcal{W}$ denote disjoint collections of chip sites and set $\mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1$. Generically, we refer to $\mathcal{C}$ as a cluster. Given $\hat{p} = N_1 / W$ and $\mathcal{C}$, the remaining $N_1 - C_1$ defective chip sites are selected without replacement from the set $\mathcal{W} \backslash \mathcal{C}$. As in Section 4.2, it is possible to compute the exact first and second moments of $T_{\mathcal{N}_1}$ and $T_{\mathcal{N}_0}$ conditionally on the event that the chips in $\mathcal{C}_0$ are good and the chips in $\mathcal{C}_1$ are defective:

$$\begin{aligned} E\left(T_{\mathcal{N}_1} \mid \hat{p} = \frac{N_1}{W},\, \mathcal{C}\right) &= T(\mathcal{C}_1, \mathcal{C}_1) + [T(\mathcal{C}_1, \mathcal{C}^c) + T(\mathcal{C}^c, \mathcal{C}_1)]\, \frac{N_1 - C_1}{W - C} \\ &\quad + T(\mathcal{C}^c, \mathcal{C}^c)\, \frac{[N_1 - C_1]\,[N_1 - C_1 - 1]}{[W - C]\,[W - C - 1]}\,, \end{aligned}$$

and

$$\begin{aligned} E\left(T_{\mathcal{N}_0} \mid \hat{p} = \frac{N_1}{W},\, \mathcal{C}\right) &= T(\mathcal{C}_0, \mathcal{C}_0) + [T(\mathcal{C}_0, \mathcal{C}^c) + T(\mathcal{C}^c, \mathcal{C}_0)]\, \frac{N_1 - C_0}{W - C} \\ &\quad + T(\mathcal{C}^c, \mathcal{C}^c)\, \frac{[N_1 - C_0]\,[N_1 - C_0 - 1]}{[W - C]\,[W - C - 1]}\,, \end{aligned}$$

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 $T(\mathcal{N}_1, \mathcal{N}_1)$ and $T(\mathcal{N}_0, \mathcal{N}_0)$ are simply $T_{\mathcal{N}_1}$ and $T_{\mathcal{N}_0}$, respectively. 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 $\mathcal{C}$. By again appealing to the theory of U-statistics (see Lee, 1990), it can be shown that the joint distribution of the pair $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$ is approximately bivariate normal provided the difference between $W$ and $C$ is not small.

In Figures 7 and 8 we make use of these approximations to obtain probability contours for our test statistics under three different clustered alternatives, each with $\mathcal{C}_0$ taken to be the empty set. The ellipses are colored to match the corresponding cluster on the right, and as one moves from top to bottom, the clusters contain 40%, 30% and 20% of the chips, respectively. To hold the yield of the wafers fixed we must decrease the background noise from 20% to 40% as indicated in the figures. The wafers in Figure 7 contain 250 chips while those in Figure 8 are comprised of 500 chips. Observe that as the background noise increases and the size of the cluster decreases, the ellipses tend toward the approximate probability contour calculated under CSR.

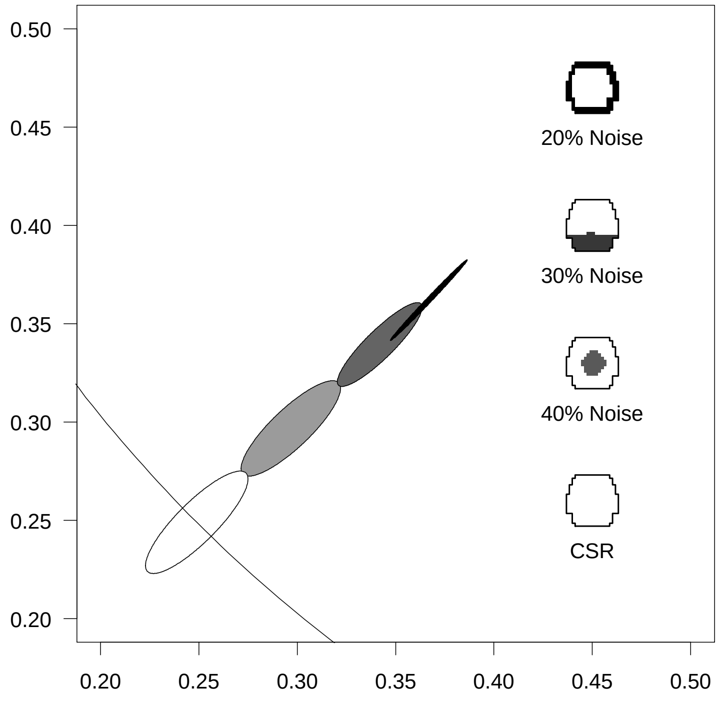


**Figure 7.** *Approximate probability contours for the pair* $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$ *calculated under the null hypothesis of spatial homogeneity and three clustered alternatives (250 chips/wafer).*

In the previous section, we calculated the mean of the pair $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$ under the null hypothesis of CSR. Using the expressions for the conditional means of $T_{\mathcal{N}_1}$ and $T_{\mathcal{N}_0}$ given above, we can also obtain the mean of our test statistics under the three clustered alternatives in Figures 7 and 8. As should be clear from the figures, the slope of the line connecting these two sets of means is essentially one. In general, it can be shown that the slope of this line is given by

$$\frac{\Delta_0 + \delta_0 - T(\mathcal{C}_0, \mathcal{C}_1)}{\Delta_0 + \delta_1 - T(\mathcal{C}_1, \mathcal{C}_0)},$$

where

$$\Delta_0 = \frac{C_0}{W} + T(\mathcal{C}^c, \mathcal{C}^c)\left(\frac{N_0 - C_0}{W - C}\right)^2 - \left(\frac{N_0}{W}\right)^2,$$

and

$$\delta_i = \left(T(\mathcal{C}^c, \mathcal{C}_i)\frac{N_i - C_i}{W - C} - T(\mathcal{C}_i, \mathcal{C}^c)\frac{N_{1-i} - C_{1-i}}{W - C}\right) - T(\mathcal{C}^c, \mathcal{C})\left(\frac{N_i - C_i}{W - C}\right)^2,$$

for $i = 0, 1$. If our cluster $\mathcal{C}$ is in fact a cluster (a spatially contiguous region) and not a disconnected collection of chip sites, then $T(\mathcal{C}, \mathcal{C}^c)$ is likely to be small. Similarly, the terms $T(\mathcal{C}_i, \mathcal{C}^c)$ and $T(\mathcal{C}^c, \mathcal{C}_i)$, $i = 0, 1$, are also likely to be small. Therefore, whether or not the slope in question is near one depends only on $T(\mathcal{C}_1, \mathcal{C}_0)$ and $T(\mathcal{C}_0, \mathcal{C}_1)$, measures of how *mixed* the $\mathcal{C}_0$ and $\mathcal{C}_1$ components of $\mathcal{C}$ are.

Now, recalling the principal axis results of the previous section, we conclude in light of Figures 7 and 8 that the sum $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ is the best linear combination of our two test statistics for detecting clustered departures from CSR. This quantity is equivalent to the statistic $BW$ discussed by Cliff and Ord (1981), which they recommend on the basis of asymptotic relative efficiency arguments. As mentioned earlier, our intuitive graphical argument also indicates that as $\hat{p} = N_1 / W$ tends to one or zero, the slope of the principal axis tends to zero or infinity, respectively, and that in these extreme situations, a different linear combination of $T_{\mathcal{N}_1}$ and $T_{\mathcal{N}_0}$ outperforms the simple sum. However, as we have already noted, the issue of spatial clustering is practically important only when the yield is is not too low and not extremely high; and in this region, the sum $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ performs well. Thus, we will focus the remainder of our discussion on the statistic $T_{\mathcal{N}_0} + T_{\mathcal{N}_1}$

## 4.3 The null hypothesis of a Markov random field

While the notion of complete spatial randomness has considerable theoretical appeal, it is not a realistic assumption from a practical standpoint. We have found that a reasonable null distribution should allow for some form of mild spatial clustering. To accommodate this change, our somewhat platonic view of the two major defect mechanisms presented

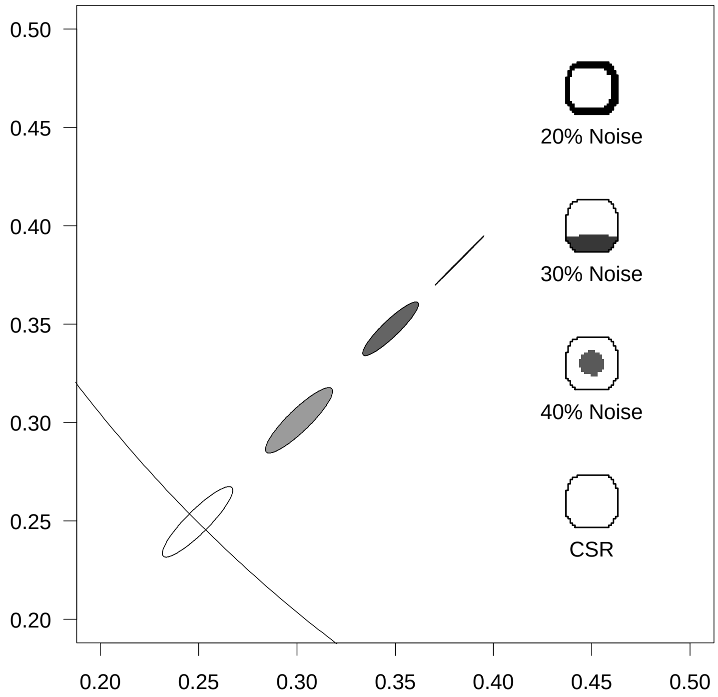


**Figure 8.** *Approximate probability contours for the pair* $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$ *calculated under the null hypothesis of spatial homogeneity and three clustered alternatives (500 chips/wafer).*

in Figure 2 should now be interpreted as a contrast between small-scale clustering and the presence of large-area defects. In this section, we discuss modeling the small-scale effects by a suitably parameterized Markov random field (MRF). While the MRF framework has historically been used in the context of hypothesis testing as a convenient class of alternatives to CSR, we are interested in these models as a tool for generating more realistic null distributions.

Under some mild conditions on our choice of weights $\{w_i(\cdot)\}$, the probability measure that assigns to the vector

$$\mathbf{I}_{\mathcal{N}_1} = \big(I_{\mathcal{N}_1}(1), \ldots, I_{\mathcal{N}_1}(W)\big)$$

the probability

$$P(\mathbf{I}_{\mathcal{N}_1}) \;=\; C\exp\left(\beta_0\frac{N_1}{W} + \beta_1(T_{\mathcal{N}_1} + T_{\mathcal{N}_0})\right), \tag{4.1}$$

is a MRF (see Strauss, 1977 and Cressie, 1991). Because $\mathbf{I}_{\mathcal{N}_1}$ is a binary vector, this distribution is also known as an auto-logistic model (see Besag, 1974). Under the specification in (4.1), we find that for any site $i \in \mathcal{W}$,

$$\log\frac{P\left(I_{\mathcal{N}_1}(i) = 1 \mid I_{\mathcal{N}_1}(j), j \neq i\right)}{P\left(I_{\mathcal{N}_1}(i) = 0 \mid I_{\mathcal{N}_1}(j), j \neq i\right)} \;=\; \beta_0 + \beta_1\sum_{j\in\mathcal{W}}\left(w_i(j) + w_j(i)\right)\left(I_{\mathcal{N}_1}(j) - I_{\mathcal{N}_0}(j)\right),$$

so that by setting $\beta_1$ equal to zero, we return to a spatially homogeneous model with

$$p \;=\; EI_{\mathcal{N}_1}(i) \;=\; \frac{\exp(\beta_0)}{1+\exp(\beta_0)}.$$

More refined comparisons between this model and the null distribution of CSR require extending theoretical results derived by Kryscio et al. (1980). A straightforward extension of their results indicates that, given $\hat{p} = N_1/W$, the sum $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ is approximately normally distributed for large values of $W$. If we let $\mu_{CSR}$ and $\sigma^2_{CSR}$ denote the conditional mean and variance of $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ given $\hat{p}$, as calculated in Section 4.1 under the hypothesis of CSR, then under the Markov random field model (4.1), the statistic $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ is again normally distributed but with mean $\mu_{CSR} + \beta_1\sigma^2_{CSR}$ and variance $\sigma^2_{CSR}$. Therefore, tests based on the null distribution (4.1) differ from those obtained under a spatially homogeneous model by a simple shift that depends only on the value of the parameter $\beta_1$. These results can be established easily following the arguments of Kryscio et al. (1980).

By setting $w_i(j) = 1$ for all $j$ such that $|i-j| = 1$, we obtain the two-dimensional Ising model from statistical mechanics. A procedure for simulating from this distribution conditionally on the value of $\hat{p}$ was developed by Flinn (1974) and was later adapted by Cross and Jain (1983) to generate images of textures that are modeled by MRF's. Using Flinn's algorithm and specifying yields of 20%, 40%, 60% and 80%, we produced the three sets of wafers displayed in Figure 9. Readers might find it interesting to compare these three sets of wafers with those in Figure 5 which were generated under the CSR model. Plots such as these can be used to determine a level of "acceptable" clustering, or equivalently, a value of the clustering parameter $\beta_1$ in equation (4.1). Finally, while we have focused our discussion on the quantity $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$, it is possible to treat the two statistics separately. In Figure 10, we plot the conditional mean of $T_{\mathcal{N}_1}$ versus the conditional mean of $T_{\mathcal{N}_0}$ given $\hat{p}$ for various values of $\beta_1$. As was the case with the deviations from CSR examined in the

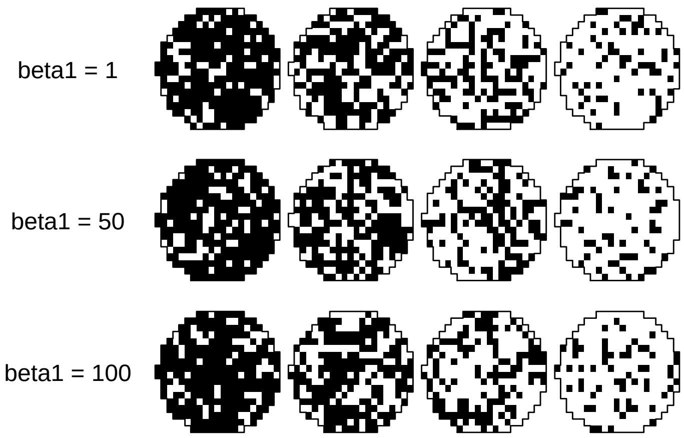


**Figure 9.** *Samples of wafers simulated under the hypothesis of an MRF. The wafers in each row have yields 20%, 40%, 60% and 80%, respectively (300 chips/wafer). These maps should be compared to the ones appearing in Figure 5.*

previous section, as the level of clustering increases, the mean of these two statistics follows a line with unit slope.

We can now describe the distribution of $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ when the superposition model in Figure 2 combines a Markov random field with a significant clustered defect. By once again appealing to Kryscio et al. (1980), we can determine the distribution of $T_{\mathcal{N}_1}+T_{\mathcal{N}_0}$ conditional on the placement of the cluster $\mathcal{C}$ and the overall proportion of defectives $\hat{p} = N_1 \,/\, W$. Using an argument similar to the one given above, we find that if $\mu$ and $\sigma^2$ denote the mean and variance of $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ conditional on $\hat{p}$ and $\mathcal{C}$ as calculated in Section 4.2, then under this new model, the sum $T_{\mathcal{N}_1}+T_{\mathcal{N}_0}$ has an approximately normal distribution with mean $\mu+\beta_1\sigma^2$ and variance $\sigma^2$ (again conditionally on $\hat{p}$ and $\mathcal{C}$). As was the case with using (4.1) as a null distribution, the only essential difference in allowing for mild clustering as the background defect process in our superposition model is the need to shift the mean of our test statistic by an amount that is known up to the constant $\beta_1$.

## 4.4 Distribution under a nonhomogeneous Bernoulli process

Many steps in the IC fabrication process possess a circular symmetry that often results in the defect probabilities $p_i = EI_{\mathcal{N}_1}(i)$, $i \in \mathcal{W}$, being higher for edge chips. Excessive wafer

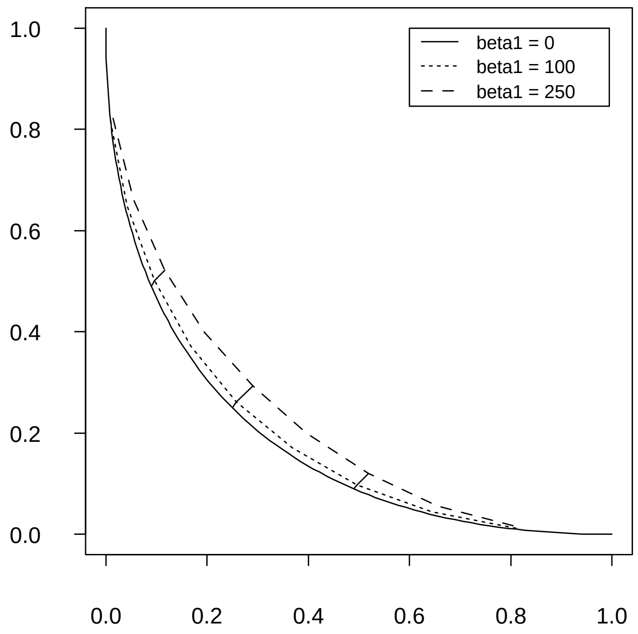


**Figure 10.** *Plot of* $E\left(T_{\mathcal{N}_1}|\hat{p}\right)$ *versus* $E\left(T_{\mathcal{N}_0}|\hat{p}\right)$ *for* $0 \leq \hat{p} \leq 1$ *under various MRF models. The three segments connect the conditional means calculated under MRF models with* $\beta_1$ *equal to 0, 100 and 250, respectively. Note that these lines have approximately unit slope.*

handling can also result in this radial pattern of increasing defect probabilities. Therefore, it is important that our monitoring procedure is also able to detect certain specific nonhomogeneous Bernoulli processes. With this in mind, we consider defect probabilities given by

$$\operatorname{logit}(\,\alpha p_i\,) \;=\; -\frac{1}{2}\left(\frac{\operatorname{dist}(i, \text{center})}{100/a}\right)^2 \qquad \text{for } \; i \in \mathcal{W}\,, \tag{4.2}$$

where "center" is the center of the wafer, $\operatorname{dist}(\cdot,\cdot)$ is the usual Euclidean distance between two points on the wafer, $a$ is a parameter that controls the radial variation across the wafer, and $\alpha$ is a scaling parameter. As $a$ tends to zero, the $p_i$'s, $i \in \mathcal{W}$, tend to a constant that is adjusted by $\alpha$ to control the overall yield of the wafer. In Figure 11 we present a number of maps simulated from this model as well as a greyscale version of the defect probabilities. For

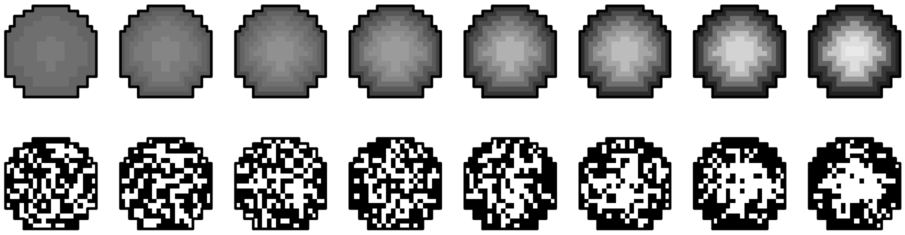

**Figure 11.** *Realizations of a nonhomogeneous Bernoulli process specified by* (4.2)*. Here a ranges from 2 to 16 in equal steps of 2 (232 chips/wafer).*

wafers in the top row of this figure, chips are shaded according to their defect probability $p_i$, with white corresponding to $p_i$ close to zero and black indicating that $p_i$ is essentially one. To make reasonable comparisons between the various models derived in this way, we scale the resulting defect probabilities so that the expected yield of each wafer is 40%. In this case, $a$ ranges from 2 to 16 in steps of 2. As expected, clusters of defective chips form around the edges of the wafer with increasing values of $a$.

This class of models can also be cast in the MRF framework discussed in Section 4.3 by setting $\beta_1$ equal to zero and replacing the ratio $N_1 \,/\, W$ with a weighted average of the variables $I_{\mathcal{N}_1}(i)$, $i \in \mathcal{W}$. The resulting measure assigns to the vector

$$\mathbf{I}_{\mathcal{N}_1} = (\, I_{\mathcal{N}_1}(1), \ldots, I_{\mathcal{N}_1}(W)\,)$$

the probability

$$P(\mathbf{I}_{\mathcal{N}_1}) \;=\; C \exp\left( \sum_{i \in \mathcal{W}} \beta_{i0} I_{\mathcal{N}_1}(i) \right) ,$$

so that

$$p_i \;=\; E I_{\mathcal{N}_1}(i) \;=\; \frac{\exp(\beta_{i0})}{1 + \exp(\beta_{i0})} .$$

Therefore, the results quoted in Section 4.3 can be used both to motivate a normal approximation for the distribution of the pair $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$, and to calculate the moments required to form probability contours for these statistics under a nonhomogeneous Bernoulli process. In this simple case, however, motivating the necessary distributional results for $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ is relatively straightforward. Essentially, under this model the triple $(N_1 \,/\, W, T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$ has an approximately trivariate normal distribution, the moments of which can be calculated

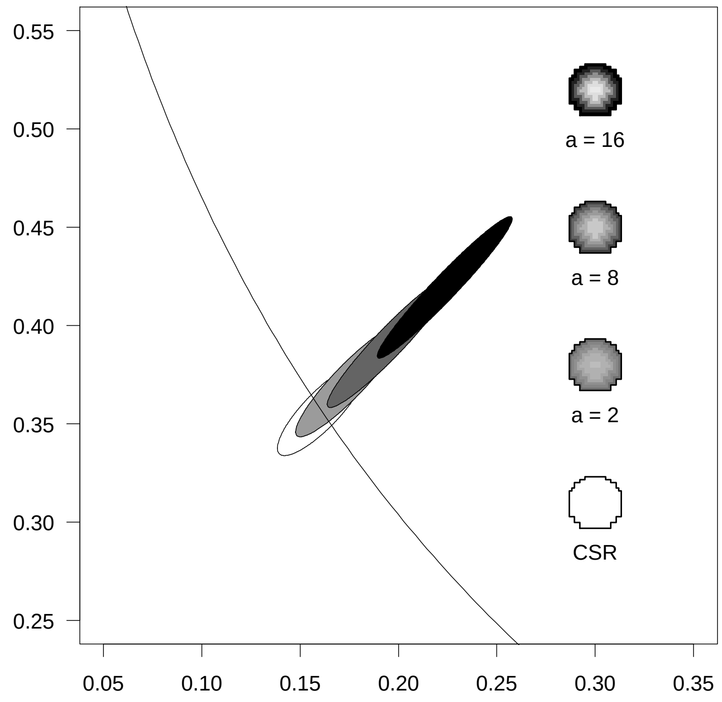


**Figure 12.** *Approximate probability contours for the pair* $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$ *calculated under the nonhomogeneous alternative specified by* (4.2) *(232 chips/wafer).*

by the accounting techniques discussed in Section 4.1. As was the case in that section, the moments depend not only on the wafer configuration $\mathcal{W}$, but also on the defect probabilities $p_i$, $i \in \mathcal{W}$, and hence must be computed for each different model we consider. By conditioning on the value of $\hat{p} = N_1 \,/\, W$, we use these calculations to obtain an approximation to the distribution of $(T_{\mathcal{N}_1}, T_{\mathcal{N}_0})$. A set of normal probability contours derived in this way corresponding to the model (4.2) are presented in Figure 12. Here, the darkest ellipse corresponds to the map with the largest value of $a$; that is, the map exhibiting the greatest degree of radial variation in the defect probabilities. Observe that once again the slopes of the lines connecting the means calculated under the null hypothesis of CSR and the various nonhomogeneous alternatives are very close to one.

## 5 The power of our monitoring procedure

In the previous section, we examined the behavior of the test statistic $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ under a variety of spatial models that mimic the patterns of defects we have observed in actual production data. Because of its relatively simple dependence on the variables $I_{\mathcal{N}_1}(i)$ and $I_{\mathcal{N}_0}(i)$, $i \in \mathcal{W}$, we were able to calculate either the exact first and second moments of this statistic under these different models or sufficiently accurate approximations. We also found that subject to certain conditions on $\mathcal{W}$, the sum $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ is essentially normally distributed with the corresponding moments. Therefore, straightforward calculations can be used to investigate the power of our test to detect various clustered alternatives at the individual wafer map level.

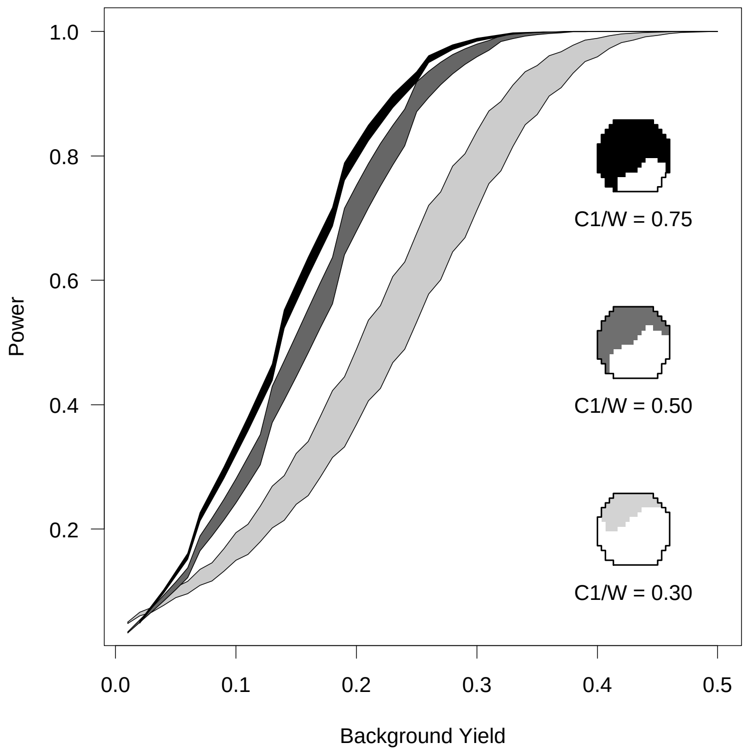


**Figure 13.** *The power that the statistic $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ has to detect several clustered alternatives assuming a MRF null distribution with values of $\beta_1$ ranging from 0 to 30 (232 chips/wafer).*

Let $\mu$ and $\sigma^2$ denote the mean and variance of the sum $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ under some spatial

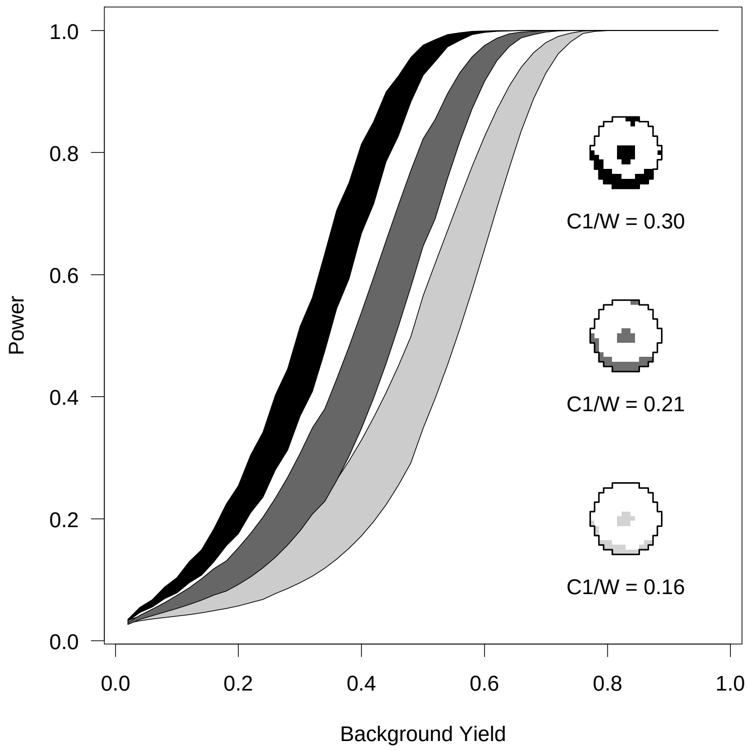


**Figure 14.** *The power that the statistic $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ has to detect several clustered alternatives assuming a MRF null distribution with values of $\beta_1$ ranging from 0 to 30 (184 chips/wafer).*

defect distribution. Given an observed wafer map, we want to test the hypothesis that $\mu \leq \mu_{CSR}$ versus $\mu > \mu_{CSR}$ conditionally on the value of $\hat{p} = N_1 / W$, where $\mu_{CSR}$ is the mean of our test statistic under a spatially homogeneous model with the indicated value of $\hat{p}$. While we focus our attention in this section on one-sided tests, there can be information in exceptionally small values of $\mu$. For example, it is intuitively clear that certain systematic patterns (say, a checkerboard arrangement of defects) will result in values of our test statistics that are smaller than would be seen under CSR. While important from a practical standpoint, for ease of presentation we will only consider testing for the presence of large-scale clusters of defects, or, equivalently, for unusually large values of $\mu$. Using the results in Sections 4.1 and 4.2, we find that the power of the $\alpha$-level test to detect clustered alternatives is reasonably approximated by

$$1 \;-\; \Phi\left(Z_{1-\alpha}\frac{\sigma_{CSR}}{\sigma} + \frac{\mu_{CSR}-\mu}{\sigma}\right) ,$$

where $Z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution. If, on the other hand, we decide instead to test against a MRF model (4.1) with a specified value of $\beta_1$, we find from Section 4.3 that the power becomes

$$1 \;-\; \Phi\left(Z_{1-\alpha}\frac{\sigma_{CSR}}{\sigma} + \frac{\mu_{CSR}+\beta_1\sigma_{CSR}^2-\mu}{\sigma}\right) .$$

For $\alpha = 0.05$, in Figures 13 and 14 we present power plotted as a function of background yield for two sets of clustered alternatives. Taking an MRF model with $\beta_1 = 0$ (CSR) as our null distribution, the power curve for the indicated alternatives follows the upper edge of each colored region. Setting $\beta_1 = 30$ (which is roughly equivalent to the value of $\beta_1$ used to simulate the second sample of wafers in Figure 9) generates the curves associated with the bottom of the three shaded areas. As $\beta_1$ runs between 0 and 30, the curves fill in the regions in the obvious way.

As one might expect, the impact of the MRF assumption is greatest for the smallest clustered defects. In any of these cases, however, if the background yield is greater than 50%, we are virtually guaranteed to detect each of the three departures from any of the null distributions considered in Figure 13. The size $C_1/\mathcal{W}$ of the smallest clustered alternative presented in Figure 13 is the same that of the largest cluster treated in Figure 14, but our testing procedure has less power in detecting the latter distribution. Intuitively, we expect this result because the cluster in Figure 14 is much more "diffuse," and the expressions derived at the end of Section 4.2 can be used to add mathematical precision to this informal reasoning. Generating the plots in Figures 13 and 14 for a variety of different clustered defects can aid in setting both the level at which we want to conduct our tests as well as the value of $\beta_1$, the clustering parameter in (4.1).

Finally, we investigate the behavior of our monitoring procedure when faced with the nonhomogeneous Bernoulli processes discussed in Section 4.4. The curve plotted in Figure 15 describes the power of our test to detect the model defined in (4.2) for various values of the parameter $a$, taking CSR as the null distribution and conditioning on the value of $\hat{p} = 60\%$. Recall that in this case, as $a$ increases, so does the degree of spatial inhomogeneity exhibited by the resulting wafer maps. In general, large values of $a$ produce clusters of defects around the edge of the wafers, and hence our power to detect these deviations from CSR increases with $a$. For ease of interpretation, we have superimposed the greyscale wafer maps of defect probabilities from Figures 11 and 12.

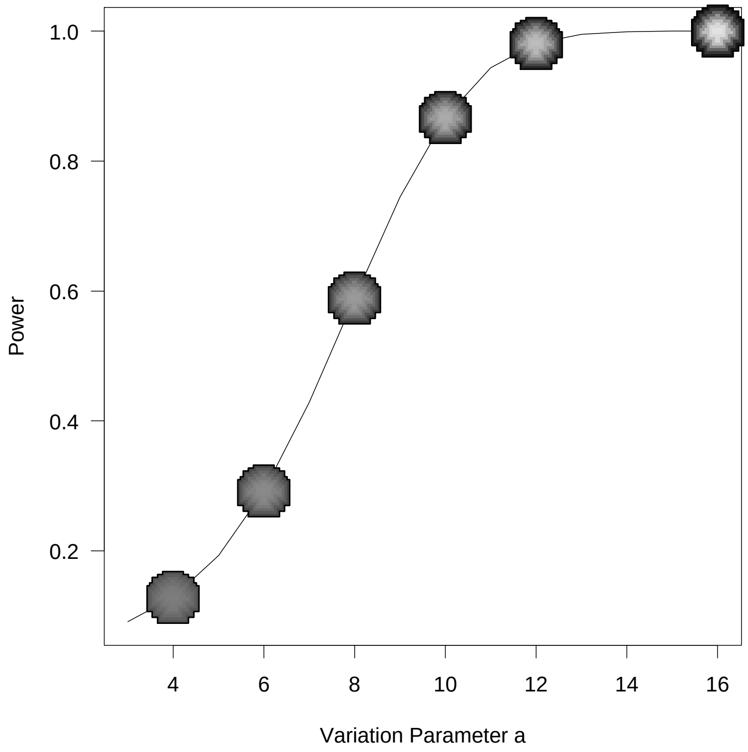


**Figure 15.** *The power that the statistic* $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ *has to detect the nonhomogeneous alternative specified by* (4.2) *(232 chips/wafer).*

We end this discussion by describing how one could use these wafer-level testing results to study the properties of a lot-level procedure for process monitoring. Specifically, we consider the automated lot-level procedure mentioned in Section 2 in which a lot is classified as exhibiting significant spatial clustering if more than some proportion $\pi$ of the wafers in the lot are identified as containing large-scale clustered defects. Let $\alpha_W$ denote the Type I error of our test of the null hypothesis of either CSR or a Markov random field model that allows for mild spatial nonhomogeneity of defects. Given $\pi$, let $\alpha_L$ denote the corresponding Type I error of this simple lot-level monitoring scheme. If the defective chips are distributed independently from wafer to wafer (a tenable assumption under the null hypothesis), then the usual binomial calculations can be used to obtain $\alpha_L$ from $\alpha_W$. If $M$, the number of wafers in the lot, is large enough to reasonably apply a normal approximation, then

$$\alpha_L \approx 1 - \Phi\left(\frac{M^{1/2}(\pi - \alpha_W)}{\sqrt{\alpha_W(1-\alpha_W)}}\right).$$

We have used this approximation to obtain the Type I errors $\alpha_L$ for $\pi = 0.2$ and for several reasonable values of $M$ and $\alpha_W$. As can be seen from Table 1, the type I errors associated with this lot-level procedure are extremely small. From the standard argument for computing the average run length, we see that $ARL_0 = 1/\alpha_L$, and hence the false alarm rate is reasonably low for these choices of $\pi$ and $\alpha_W$.

| $M$ | $\alpha_W$ | $\alpha_L$ |
|---|---|---|
| 25 | .01 | .0000 |
| 25 | .01 | .0000 |
| | .05 | .0003 |
| | .10 | .0480 |
| 40 | .01 | .0000 |
| | .05 | .0000 |
| | .10 | .0180 |

**Table 1.** *The Type I error associated with our simple lot-level screening rule calculated holding* $\pi = 0.2$.

We can also extract the power behavior of this lot-level monitoring rule from our discussion of the wafer-level hypothesis tests above. Let $\beta_W$ denote the power of the test based on $T_{\mathcal{N}_1}+T_{\mathcal{N}_0}$ to detect any of the spatially clustered alternatives presented in Section 4. If one can assume that the defects are distributed independently across wafers, then the same argument as before can be used to approximate the power of the lot-level procedure as follows:

$$\beta_L \approx 1 - \Phi\left(\frac{M^{1/2}(\pi - \beta_W)}{\sqrt{\beta_W(1-\beta_W)}}\right).$$

Using this approximation, we calculated $\beta_L$ for $\pi = 0.2$ and for several reasonable values of $M$ and $\beta_W$. These values, which are shown in Table 2, indicate that the procedure has good

| $M$ | $\beta_W$ | $\beta_L$ |
|---|---|---|
| 25 | .90 | 1.0000 |
| | .75 | 1.0000 |
| | .50 | 0.9987 |
| | .25 | 0.7180 |
| 40 | .90 | 1.0000 |
| | .75 | 1.0000 |
| | .50 | 0.9999 |
| | .25 | 0.7670 |

**Table 2.** *The power of our simple lot-level screening rule calculated holding* $\pi = 0.2$.

power in detecting lots with spatially clustered defects. The values in Table 2 were based on the assumption that the defective chips are distributed independently from wafer to wafer. In practice, we expect that many wafers in the lot might experience the same processing problems and hence share common patterns of defects. In the presence of such dependence, the $\beta_L$ can vary from the value listed in the above table (when the defects are completely independent across wafers) to $\beta_W$ (when all the wafers exhibit the same problem). Thus, $\beta_W$ provides a lower bound to the power of the lot-level procedure. Our overall strategy for process improvement explicitly recognizes the fact that the clustered defects in a lot may be caused by one or more process problems and hence the wafer maps will exhibit similar spatial patterns. We have used this fact to create graphical displays and pattern recognition techniques to further enhance the effectiveness of the wafer-level and lot-level monitoring procedures outlined above. The interested reader is referred to Friedman, et al. (1993, 1995, 1996) and Hansen and Nair (1995).

## 6 Concluding remarks

Our focus in this paper has been on methods for routinely monitoring data collected at final probe to detect the presence of spatially clustered defects. We have emphasized wafer-level tests based on the statistic $T_{\mathcal{N}_1} + T_{\mathcal{N}_0}$ because of the tremendous wafer-to-wafer variability that we have observed in our environment. This appears to be common in applications-specific low volume production. Clearly, this may not be the case at facilities manufacturing large volumes of a single code, and one might prefer to directly monitor lot-level composite (sitewise average) wafer map data. Extensions of both the theory and the methodology presented here to composite wafer maps will be discussed elsewhere.

Monitoring procedures such as these are just the first step in an effort to identify potential problems in the fabrication process. As alluded to at the end of the previous section, we have developed a comprehensive statistical approach for quality and and process improvement that is built in part on the material developed in this article. Our methods have been applied within the IC manufacturing division of AT&T for the past four years and have led to significant process and yield improvements. Throughout our involvement in this area, we have emphasized the importance maintaining the spatial character of the data collected during the IC fabrication process. We have developed techniques for wafer visualization and the detection of defect signatures and have created a variety of graphical tools for relating various types of wafer-level measurements with final probe data. S-based software that implements these techniques has been developed for UNIX workstations and PC's.

## 7 Acknowledgements

This work has been done in collaboration with the AT&T Microelectronics facility in Orlando. The authors are grateful to J. R. Hoyer for his assistance in formulating the problem, helpful discussions and insightful comments. We also wish to express our thanks to D. James who has been a significant part of our interaction with AT&T Microelectronics for the last three years. V. N. Nair's research was supported in part by NSF Grant DMI-9501217 and by the Center for Display Technology and Manufacturing of the University of Michigan.

# Technical Appendix

**Lemma A.1** *Let $K$ be a positive integer larger than 2, and let $I+J=K$ for $I$ and $J$ two integers that are each greater than one. Then we have the following:*

$$\frac{I[I-1]}{K[K-1]}-\frac{J[J-1]}{K[K-1]}-\left(\left[\frac{I[I-1]}{K[K-1]}\right]^2-\left[\frac{J[J-1]}{K[K-1]}\right]^2\right)$$
$$= \quad 2\left[\frac{I}{J-1}-\frac{J}{I-1}\right]\frac{I[I-1]}{K[K-1]}\frac{J[J-1]}{K[K-1]}$$

$$\frac{I[I-1][I-2]}{K[K-1][K-2]}-\frac{J[J-1][J-2]}{K[K-1][K-2]}-\left(\left[\frac{I[I-1]}{K[K-1]}\right]^2-\left[\frac{J[J-1]}{K[K-1]}\right]^2\right)$$
$$= \quad \frac{K-4}{K-2}\left[\frac{I}{J-1}-\frac{J}{I-1}\right]\frac{I[I-1]}{K[K-1]}\frac{J[J-1]}{K[K-1]}$$

$$\frac{I[I-1][I-2][I-3]}{K[K-1][K-2][K-3]}-\frac{J[J-1][J-2][J-3]}{K[K-1][K-2][K-3]}-\left(\left[\frac{I[I-1]}{K[K-1]}\right]^2-\left[\frac{J[J-1]}{K[K-1]}\right]^2\right)$$
$$= \quad \frac{-4}{K-2}\left[\frac{I}{J-1}-\frac{J}{I-1}\right]\frac{I[I-1]}{K[K-1]}\frac{J[J-1]}{K[K-1]}$$

**Proof:** Algebra.

**Proposition A.2** *Set*

$$\sigma = \operatorname{sign}\left(\alpha_4\frac{W-3}{W-1}-\frac{W-2}{W}\right).$$

*Then, the principal component of the variance-covariance matrix of $(T_{\mathcal{N}_0},\, T_{\mathcal{N}_1})$ is given by the vector*

$$\begin{pmatrix} 1 \\ \frac{\sigma\sqrt{4+\gamma^2}-\gamma}{2} \end{pmatrix},$$

*where*

$$\gamma \;=\; \left(\frac{(2\alpha_2+\alpha_3)-\frac{4}{W}}{\alpha_4\frac{W-1}{W-3}-\frac{W-2}{W}}\right)\left[\frac{W-D}{D-1}-\frac{D}{W-D-1}\right] \;\approx\; C\left[\frac{\hat{q}}{\hat{p}}-\frac{\hat{p}}{\hat{q}}\right]$$

*and $C$ is a constant that depends only on $\mathcal{W}$ and our choice of weights $\{w_i(\cdot)\}$.*

**Proof:** We recall from standard multivariate analysis that if our variance-covariance matrix is given by

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix},$$

then the eigenvector corresponding to the largest eigenvalue is given by the vector $(1, \eta)$, where

$$\eta = \frac{\sigma_2^2 - \sigma_1^2}{2\sigma_{12}} + \sqrt{\left(\frac{\sigma_2^2 - \sigma_1^2}{2\sigma_{12}}\right)^2 + 1}.$$

The proposition follows by substituting the appropriate values for $\sigma_1^2$, $\sigma_2^2$, and $\sigma_{12}$ using the moment results quoted in Section 2 along with a straightforward application of Lemma A.1.

## References

1. Albin, S. L. and Friedman, D. J. (1989). "The impact of Clustered Defect Distributions in IC Fabrication," *Management Science* **35** 1066-1078.

2. Besag, J. (1974). "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *JRSSB* **36** 192-236.

3. Cliff, A. D., and Ord, J. K. (1981). Spatial Processes, Models and Applications, Pion Limited, London.

4. Cressie, N. (1990). Statistics for Spatial Data, John Wiley & Sons, New York.

5. Cross, G. and Jain, A. (1983). "Markov random field texture models," *IEEE Trans. PAMI* **5** 25-39.

6. Cunningham, J.A. (1990). "The Use and Evaluation of Yield Models in Integrated Circuit Manufacturing," *IEEE Trans. Semicond. Manufact.* **3** 60-71.

7. Flinn, P. A. (1974). "Monte Carlo calculation of phase separation in a two-dimensional Ising system." *J. Stat. Phys.* **10** 89-97.

8. Friedman. D. J., Hansen, M.H. Hoyer, J.R and Nair, V.N., (1993), "A Method for Characterizing Defects in Integrated Circuits" U.S. Patent No. 5,240,866.

9. Friedman, D. J., Hansen, M. H., Nair, V. N., and James, D. A. (1995). "Model-Free Estimation of Some Yield Metrics in Integrated Circuit Fabrication," (submitted, *IEEE Trans. Semicond. Manufact.*).

10. Friedman, D. J., Hansen, M. H., and Nair, V. N. (1996). "Yield and Process Improvement in Integrated Circuit Fabrication Using Spatial Information from Wafer Map Data," (in preparation).

11. Hansen, M. H. and Nair, V. N. (1995). "Process and Productivity Improvement in Semiconductor Manufacturing," in *Computing Science and Statistics* **27** (Meyer, M. and Rosenberger, J. eds).

12. Kryscio, R., Saunders, R. and Funk, G. (1980). "Normal approximations for binary lattice systems," *J. Appl. Prob.* **17** 674-685.

13. Lee, A. J. (1990). $U$-Statistics, Theory and Practice, Marcel Dekker, New York.

14. Mallory, C.L., D.S. Perloff, T.F. Hasan, and R.M. Stanley (1983). "Spatial Yield Analysis in Integrated Circuit Manufacture," *Solid State Technology* **26** 121-127.

15. Moran, P. A. P. (1948). "The interpretation of Statistical Maps." *JRSSB* **10** 243-251.

16. Strauss, D. (1977). "Clustering on coloured lattices," *J. Appl. Prob.* **14** 135-143.

17. Taam, W. and Hamada, M. (1993). "Detecting spatial effects from factorial experiments: an application in integrated-circuit manufacturing," *Technometrics* **35** 149-160.