"Statistical Process Monitoring of Correlated Binary and Count Data Using Mixture Distributions"

Brenda Cantell, Randy Collica, and José Ramírez
Digital Semiconductor, Hudson, MA

ABSTRACT

Statistical Process Monitoring (SPM) is used quite extensively in semiconductor manufacturing, particularly in the areas of yield enhancement and contamination (particles) reduction. These variables are typically measured at the Integrated Circuit (IC)-level, and then summarized to the wafer and lot levels. Due to the batch nature of semiconductor processing, ICs residing in the same wafer are normally processed simultaneously. This tends to induce a positive correlation among the measurements taken at the IC level.

It is known that within subgroup correlation affects the performance of control charts. In particular, it tends to induce higher than expected false alarm rates. In this paper we present 3-sigma Shewhart charts, for both yield and particle data, and CUSUM monitoring charts for yield data. These charts use mixture distributions to take into account the inherent correlation in the data; thus reducing the false alarm rate.

A SAS/AF® application has been developed to deploy the use of our SPM schemes. The application allows the engineers and analysts to seamlessly: (1) estimate the parameters of these mixture distributions (beta-binomial for yield data and negative binomial for particle data) using SAS/IML® optimization routines, and (2) produce the corresponding customized control charts using SAS/GPLOT® and SAS/SHEWHART®.

Key Words: Beta-binomial, Binomial, CUSUM, Negative binomial, Overdispersion, Poisson, Shewhart Chart.

INTRODUCTION

Semiconductor manufacturing of integrated circuits is a very complex process consisting of hundreds of processing steps. With the leaps and bounds in technology, e.g., integrated circuits with 3 million transistors on one chip using 0.35 μm width lines, implementing reliable Statistical Process Monitoring (SPM) approaches are becoming more and more challenging. SPM is applied to yield enhancement and defect or particles reduction. In this paper we will discuss appropriate SPM approaches for these two responses.

Yield is used, in part, to measure the quality of wafer fabrication processes. This measurement is taken on each individual IC on a wafer that consists of *k* ICs, and is represented by one of two values. For example, if the IC is functional then yield=1, otherwise yield=0. The number of good ICs per wafer or per lot (a grouping of 24 wafers) is often used for analyses. More specifically, wafer yield and lot yield are rigorously assessed for trends or any type of predictable patterns that might result in a long term yield gain or loss.

In-line particulate contamination can seriously affect the yields of the ICs. This is one reason why state-of-the-art clean room facilities have become as important as state-of-the-art wafer fabrication processing. A significant amount of resources are utilized to measure, monitor, and control particulate contamination during wafer fabrication. Statistical process monitoring of particle counts on wafers, in the cleanroom air, or in the chambers of equipment is a necessary part of quality control.

Since both particle counts and yield data are attribute variables, attribute based control charts should be considered for monitoring purposes. Wafer and lot yield are composed of the sum of binary variables (functional IC or non-functional IC), which results in a binomial random variable, while particle counts, in space or time, can be described by a Poisson process. In the statistical process control literature, np or p charts are used with binomial random variables, and c or u charts are used with Poisson random variables.

Control charts do not always perform the way that they are portrayed in textbooks. Violations in the assumptions of the underlying models are one reason for this. In the remainder of this paper, we describe ways that semiconductor yield and particle data may violate the assumptions of p and u charts, respectively. We also present alternative ways of developing SPM strategies using mixture distributions to account for specific violations in model assumptions. A SAS/AF® application has been developed to deploy the use of our SPM schemes. This allows the engineers and analysts to seamlessly: (1) estimate the parameters of these mixture distributions (beta-binomial for yield data and negative binomial for particle data) using SAS/IML® optimization routines, and (2) produce the corresponding customized control charts using SAS/GPLOT® and SAS/SHEWHART®.

WHY ATTRIBUTE CHARTS FAIL

All control charts make an assumption about the underlying distribution of the variable that is being charted, and the correlation structure (or lack thereof) within the subgroup. For example, np and p charts assume that the binomial distribution adequately describes the charting variable, while the Poisson distribution provides the basis for c and u charts. This is relevant because the control limits for these charts are derived using an estimate of the variance from the underlying distributions. For example, the control limits for a np chart will be derived from the binomial variance $np(1-p)$.

In semiconductor manufacturing, the subgroup observations for wafer (lot) level yield or particle data are typically not independent from each other. For instance, individual IC yield on a wafer are correlated to each other and may result in clustered patterns of good or bad ICs on a wafer. This clustering phenomenon has been studied and is well documented in the literature (See for example, Cliff and Ord (1981) or Stapper (1986)). However, its impact on statistical monitoring schemes is not always discussed.

The clustering or correlation among subgroup measurements induces a violation in the basic assumption of independent subgroup samples; which in turn will effect the variance that is used to calculate the control limits. If the correlation between the measurements in the subgroup is positive, the data will exhibit a variance that is larger than the one predicted by the specified model; a phenomenon called *overdispersion*. For example, if we consider a binomial model for yield data the variance is given by $np(1-p)$. However, when positive correlation is present (see Ramírez and Cantell (1997)) the variance is given by $Q \times np(1-p)$, with $Q > 0$. In other words, the yield variance will be larger, by a factor of Q, than is expected under a binomial model. The quantity Q is the overdispersion factor that needs to be estimated from the data.

Not accounting for overdispersion can have a detrimental impact on the control limits of attribute based control charts. Typically, the control limits will be unusually tight and will result in a higher rate of false alarms. Wheeler and Chambers (1992), page 271, notice that "when there is clustering [...], the charts will almost automatically be out of control".

Figure 1a shows a p-chart for lot yield data generated using PROC SHEWHART with the PCHART statement. This chart assumes a binomial distribution, and does not account for overdispersion. The chart shows very tight control limits, which result in a large number of out-of-control points. Many of these points are not associated with assignable causes of variation and are considered false alarms.

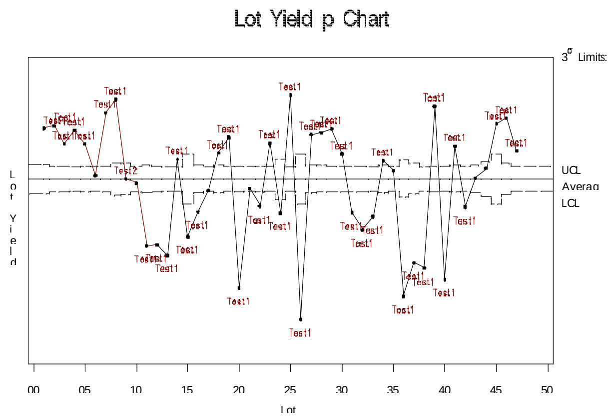


Figure 1a. P-Chart for 47 lots showing 40 points out of control.

The same phenomenon can occur with count data. Figure 1b shows a u-chart for particle data per unit time, that was collected using a sensor located on the inside of a chamber of a piece of equipment. The chart was created using PROC SHEWHART with the UCHART statement. Once again, there are a high number of false alarms due to unusually tight control limits.

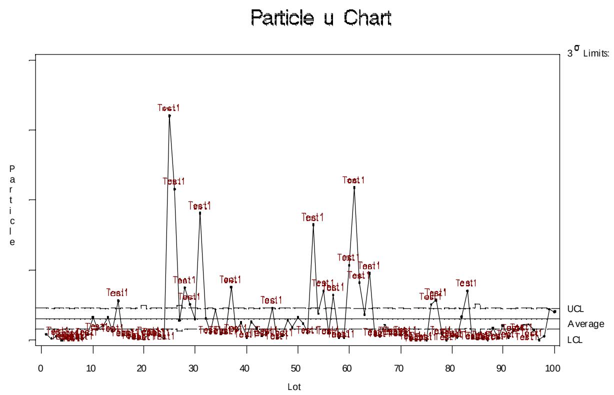


Figure 1b. U-Chart for 100 points showing 75 points out of control.

It should be evident that standard p and u control charts are unreliable to monitor yield and particle data that exhibit correlation. It is our experience that correlation is a common phenomenon in semiconductor manufacturing. More appropriate statistical methods for monitoring yield and particle data are described next.

PROCESS MONITORING OF YIELD DATA

In this section, we describe statistical process monitoring schemes for both large (Shewhart) and small (cumulative sum) shift detection for the mean yield using the beta-binomial distribution.

Beta-Binomial 3-Sigma Shewhart Charts

It is typically assumed that the sum of binary random variables, e.g. wafer yield, the yields of all of the integrated circuits in the wafer, can be modeled using the binomial distribution. This, however, assumes independence between the binary random variables. In other words, we assume that the likelihood of an integrated circuit being functional will not depend on its neighbor being either functional or not functional. Ramírez and Cantell (1997) discuss why this assumption does not hold in semiconductor manufacturing due to the "wafer effect", or the clustering of good and bad integrated circuits within a wafer. This clustering of ICs generally creates a positive correlation among the IC yields, which in turn increases the variance of the wafer yield.

One way of modeling this phenomenon is to use a mixture distribution that takes into account both the binary nature of the data, and the clustering of ICs within a wafer. Ramírez and Cantell (1997) recommend the use of the beta-binomial distribution. In this approach we assume that the common within wafer correlation, ρ, is a function of the shape parameters, α and β, of a beta distribution; i.e., $\rho = 1/(1+\alpha+\beta)$, and that the wafer yields follow a beta-binomial distribution. The parameters of the beta-binomial can be estimated by maximum likelihood or by the method of moments.

For each lot (subgroup) of wafers in our sample we use a beta-binomial distribution to estimate the mean $P = \alpha/(\alpha + \beta)$, and variance $\sigma^2 = P(1-P)/(\alpha+\beta+1)$ for the subgroup. Maximum likelihood estimation, using quasi-newton methods, is performed using a PROC IML macro. The lot means are plotted in the chart while the lot standard deviations are used to calculate an overall estimate of the process standard deviation. Since the number of wafers per lot varies, i.e., the subgroup size is not constant, we need an estimate of the process variability that takes this into account.

The chart can be easily generated using the XCHART statement and the TABLE option in PROC SHEWHART. The TABLE option allows the user to create a chart from the beta-binomial lot summary statistics. The overall estimate of process variability can be computed using the SMETHOD = MVLUE option. The control limits are given by the mean of the lot means $\pm$ 3 sigma, where sigma is the overall estimate of process variability.

Figure 2 shows a Shewhart chart with 3σ limits based on the MVLUE estimate of the lot beta-binomial standard deviations. Notice that only one lot is below the lower control limit, indicating a low yield situation. This lot was associated with nonstandard processing. This is in sharp contrast with the p-chart in Figure 1a, in which 40 out of 47 lots were either above or below the control limits!

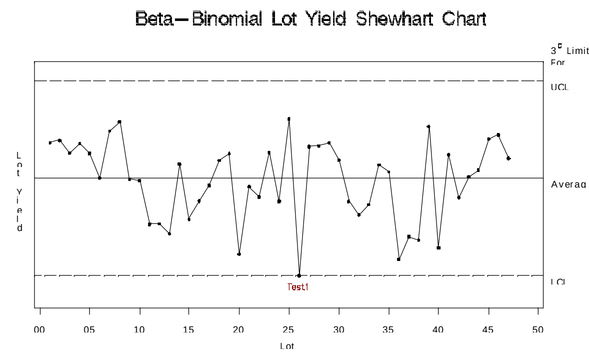


Figure 2. Shewhart Chart with Beta-binomial based limits. Only one point out of control.

By taking into account the "wafer-effect", i.e., the clustering of integrated circuits within a wafer, the number of false alarms has been reduced by more than 80%.

The overdispersion parameter Q, being a function of the within wafer correlation or wafer-to-wafer variation, can be used to monitor the within lot variation. Since Q is distributed as a Chi-Squared (see Cantell (1993)), the limits for this type of chart are calculated using this distribution.

Figure 3 shows an overdispersion chart produced using the SCHART statement in PROC SHEWHART. Once again we use the TABLE option to work with the Q estimates for each lot. Even though most of the lot yields are in control (Figure 2), we see that 8 lots have large amounts of overdispersion, while several lots fall outside the lower control limits. These latter lots are important because they are the ones with either small wafer-to-wafer variation, or small within wafer correlation; i.e., small amounts of clustering.

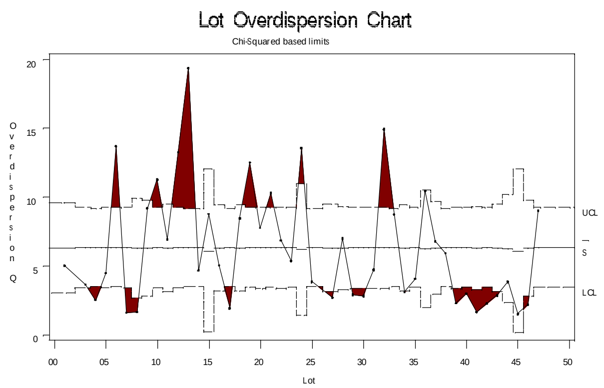


Figure 3. Overdispersion Chart with Chi-Squared based limits.

Beta-Binomial CUSUM

Even small changes in yield can result in significant cost savings. Therefore, it is desirable to monitor yield using charts that are capable of detecting small shifts in the process mean ($0.5\sigma - 2.5\sigma$). It is possible to use a p chart overlaid with Western Electric rules to increase its sensitivity to shifts. However, it is more efficient to use a chart that is designed for detecting these types of small shifts. For this application, a CUSUM (cumulative sum) chart for beta-binomial data will be employed.

CUSUM charts are typically applied to normally distributed data, and are well documented in the literature (e.g. Montgomery (1991), Ryan (1989)). Their use with attribute data is addressed less frequently, and often involves transformations of the response. Johnson and Leone (1962) give examples of CUSUM charts for Poisson and binomial distributed data that are based on Wald's sequential probability ratio tests (SPRT). Using Wald's approach, it is possible to develop a CUSUM chart for yield data using the beta-binomial distribution.

For a beta-binomial distribution with parameters α_0 and β, the population mean is given by $P_0 = \alpha_0/(\alpha_0 + \beta)$ (see for example Cox and Snell (1990)). To detect a small shift from P_0 to P_1 the SPRT is described by the following region:

$$\frac{e_2}{1-e_1} \le \prod_{i=1}^{n} \frac{\Gamma(x_i+\alpha_i)\Gamma(\alpha_1+\beta)\Gamma(k+\alpha_0+\beta)\Gamma(\alpha_0)}{\Gamma(x_i+\alpha_0)\Gamma(\alpha_0+\beta)\Gamma(k+\alpha_1+\beta)\Gamma(\alpha_1)} \le \frac{1-e_2}{e_1}$$

where e_1 and e_2 are the type I and type II error rates, respectively, and Γ is the gamma function.

This region is derived using the ratio of the density functions for the beta-binomial under $P_0 = \alpha_0/(\alpha_0 + \beta)$ and $P_1 = \alpha_1/(\alpha_1 + \beta)$. Note that the change to P_1 is induced by a change in the parameter α. However, the change in P_1 can also be induced by a change in β, or in both parameters.

The charting statistic to be plotted is the cumulative sum of an expression that involves the ratio of log gamma functions, which includes the number of good die per wafer, and the total number of die per wafer. The control limits are approximated using $\ln(e_1)$ where e_1 is the type I error rate; usually equal to $0.0027\times 5 \approx 0.01$ for 3-sigma charts.

A SAS macro was written to produce a beta-binomial CUSUM chart. The user is required to input the size of the shift, in standard deviation units, they want to detect; e.g. $\delta = 0.5$. The parameters of the beta-binomial are estimated using historical data and maximum likelihood estimation using PROC IML. Once the in-control mean ($P_0 = \alpha_0/(\alpha_0 + \beta)$) and variance ($\sigma_0^2 = P_0(1-P_0)/(\alpha_0+\beta+1)$) are calculated, the shift size is given by $P_1 = P_0 \pm \delta\times\sigma$.

We now apply this technique to the lot yield data that was presented in Figures 1(a) and 2. The target value (in-control mean) and standard deviation were estimated with historical data and δ=0.5.

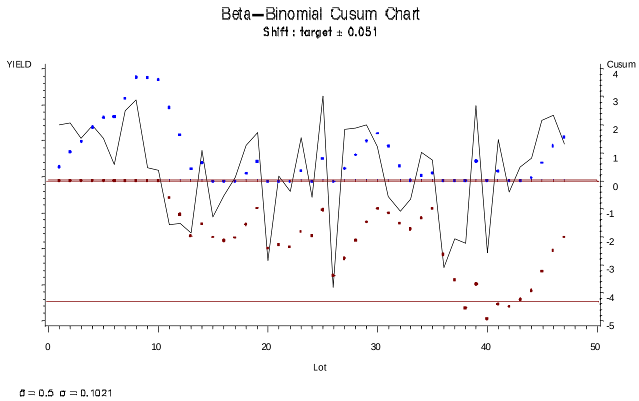


Figure 4. Beta-binomial CUSUM chart to detect a 0.5σ shift in the mean. There is an out-of-control signal at lot 38.

Figure 4 illustrates a beta-binomial CUSUM chart for lot yield data. This figure overlays two one-sided CUSUM schemes in one plot, and also shows the trend of the yield data. The left-hand axis is used to read the actual values of the yield series, while the right hand axis is used to interpret the CUSUM values. The dots that are shown above the solid line at 0 indicate upward changes in the target value. Similarly, the dots below this line indicate downward changes in the target value. The chart was designed to detect a 0.5σ shift in the in-control mean. An out-of-control signal is detected at lot 38 indicating that the mean has shifted 5% or more below the target value.

PROCESS MONITORING OF PARTICLE DATA

Particle data presents the same type of problems as yield data, i.e., overdispersion and correlated observations. This leads to narrow control limits and an excessive amount of false alarms in u and c charts, as was shown in Figure 1b.

We can use a mixture distribution to account for these phenomena. In this case, the Poisson distribution is combined with a gamma distribution to give a negative binomial distribution. The negative binomial is used to compute the limits for a 3-sigma Shewhart chart.

Once again, for each subgroup in our sample we estimate the mean and variance using a negative binomial distribution. Maximum likelihood estimation, using quasi-newton methods, is performed using a PROC IML macro. The subgroup means are plotted in the chart while the subgroup standard deviations are used to calculate an overall estimate of the process standard deviation.

The chart is produced using the XCHART statement and the TABLE option in PROC SHEWHART. The control limits are given by the mean of the subgroup means ± 3 sigma, where sigma is the overall estimate of process variability.

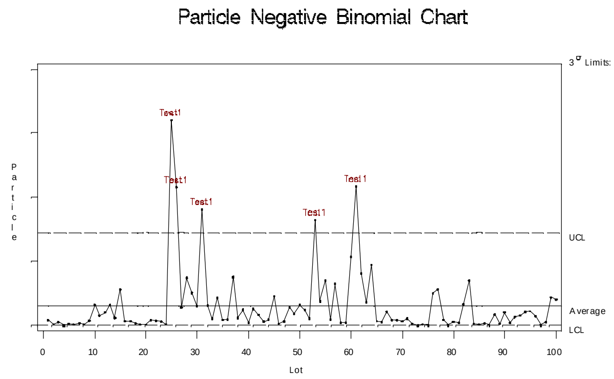


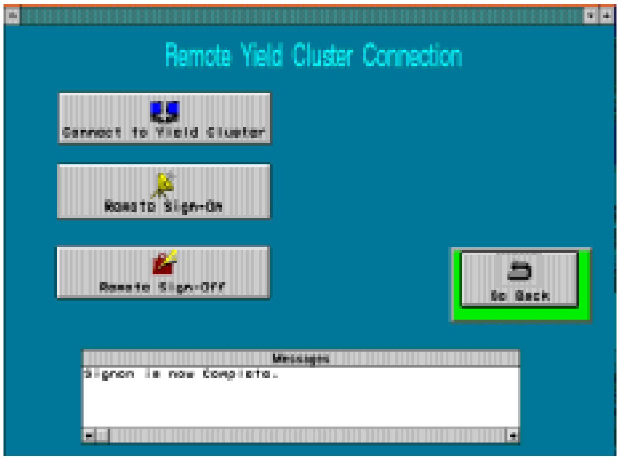
Figure 5. Shewhart chart with negative binomial based limits. Only 5 points out of control.

Figure 5 shows the resulting chart applied to the same data in Figure 1(b). Note that the number of false alarms has been drastically reduced. The five out-of-control points indicate instances where a higher than normal particle count was encountered.

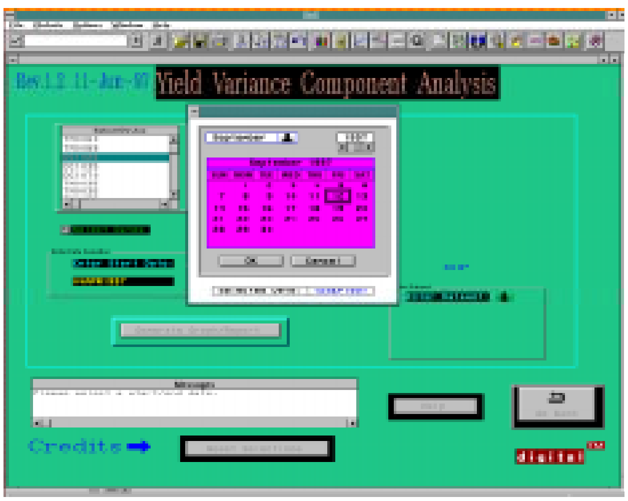
SAS GRAPHICAL USER INTERFACE

A SAS/AF® GUI interface was developed to deploy our SPM schemes for yield data to process and yield engineers. This tool has many nice features including easy data acquisition and filtering. New features in SAS® 6.12 provide the end-user with "PC" like objects; such as, a composite widget that selects dates with a calendar function which looks like a calendar, and data selection list boxes which are intuitive to the user. All statistical calculations, graphical displays, and beta-binomial control charts are transparent to the end-users. These algorithms are incorporated into the SAS/AF® program via macros.

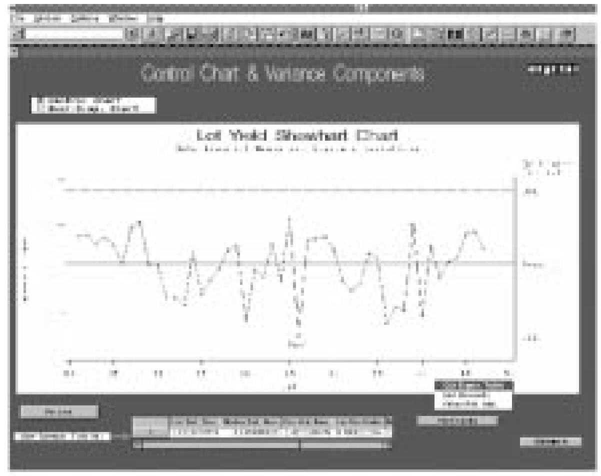
Prior to the user running the Yield SPM (YSPM) application, the user selects an icon in the SAS toolbar menu to automatically mount a virtual disk drive, run SAS/Connect®, and a sign-on process. This SAS/AF® frame is shown below.



The next SASAF® Frame, that is shown on the following page, for YSPM allows flexibility in data selection and filtering. This example illustrates how the end-users can use a calendar to segment their data using a start and end date. The date calendar is a composite widget in SAS®; originally designed by Qualex Consulting Services Inc.™ These types of features make the interface user-friendly to the end-user because they are "PC" like in nature.



The final SAS/AF® Frame that is shown below, allows the user to view the beta-binomial control chart of the data or the overdispersion chart. A variance table pull-up is available which displays the variance components numerically. A methods button allows the user to obtain other important information from the batch (lot), such as, if the lot was experimented on or if it is a production lot, remarks from the production shop-floor remarks database, and other relevant information.



CONCLUSIONS

We have shown how traditional control charting procedures can be adapted to yield and particle data. Attribute control charts that are based upon the binomial or Poisson distributions do not adequately model our data because of overdispersion. The unadjusted charts have a high false alarm rate.

Both large shift and small shift detection charts were derived using the beta-binomial and negative-binomial distributions and examples of these were presented.

These algorithms would be difficult for the average engineer or analyst to implement without the appropriate software. The SAS® system allowed us to create a "point-and-click" application that generated more appropriate charts and can be used as an exploratory tool by the end-users.

REFERENCES

1. Cantell, B.S. (1993), "Overdispersion in Binomial Models: An Application to Yield Data", unpublished thesis, Worcester Polytechnic Institute.

2. Cliff, A.D. and Ord, J.K. (1981), *Spatial Processes: Models and Applications, first edition*, Pion Ltd.

3. Cox, D.R. and Snell, E. J. (1990), *Analysis of Binary Data*, 2nd Edition, Chapman and Hall, New York, New York.

4. Johnson, N. L. and Leone, F. C. (1962), "Cumulative Sum Control Charts – Mathematical Principles Applied to Their Construction and Use –

Part III," *Industrial Quality Control*, 22 – 28.

5. Ramírez, J.G. and Cantell, B.S. (1997), "An Analysis of a Semiconductor Experiment Using Yield and Spatial Information," *Quality and Reliability Engineering International*, Vol. 13, 35-46.

6. Stapper, C.H. (1986), "On Yield, Fault Distribution, and Clustering of Particles," IBM Journal of Research and Development, Vol. 30, No. 3.

7. Stapper, C.H. (1987), "Correlation Analysis of Particle Clusters on Integrated Circuit Wafers," IBM Journal of Research and Development, Vol. 31, No. 6.

8. Wheeler, D.J. and Chambers, D.S. (1992) *Understanding Statistical Process Control*, SPC Press, Knoxville, TN.

Brenda S. Cantell, Digital Semiconductor, 79 Reed Road, Hudson, MA 01749, (978) 568-5992, brenda.cantell@digital.com.

Randall S. Collica, Digital Semiconductor, 79 Reed Road, Hudson, MA 01749, (978) 568-5119, randy.collica@digital.com.

José G. Ramírez, Digital Semiconductor, 77 Reed Road, Hudson, MA 01749, (978) 568-7400, jose.g.ramirez@digital.com.