

A Bayesian Control Chart for a Passivation Process

Balgobin Nandram

Worcester Polytechnic Institute, USA

Brenda Ramirez

W. L. Gore & Associates Inc., USA

Jai Won Choi

Center for Disease Control, USA

Multivariate control charts can be used effectively to monitor the quality of complex processes with several critical variables simultaneously. However, when the covariance matrix has large dimension in comparison to the number of runs available for parameter estimation, these charts can perform poorly. We incorporate prior information about the covariance matrix in which the number of parameters is reduced to just two. We consider a passivation process for semiconductor manufacturing, where each of the variables represents a value at a specific location in a passivation tube, and because of the interaction between the plasma and the reactant gases flowing down the tube, the correlation among the variables might decay with distance between these locations. Moreover, the variability at the locations might be taken equal, further reducing the number of parameters. We use a Bayesian method to construct the multivariate control chart, and a statistic, analogous to Hotelling's T^2 , is used for charting.

Keywords: Average Run Length, Correlation, Hotelling's T^2 , Multivariate Control Chart, Semi-conductor Data.

Introduction

Statistical process control has been applied with marginal success in the semiconductor industry. This may be, in part, due to a heavy reliance on univariate control chart practices when the quality of many production processes is characterized by a number of variables which may be highly correlated. When the data are correlated, the univariate charts are not as sensitive to out-of-control values as the multivariate charts which incorporate the correlation. It is, therefore, pertinent to use multivariate control charts to perform statistical process monitoring for such processes. However, the dimensions of the covariance matrix for the variables can expand rapidly for a complex

process, and with a small number of runs, it can be difficult to efficiently estimate the covariance matrix. We describe a new method which incorporates prior information about the parsimony of the covariance matrix. In particular, we consider applications in which the covariance matrix is reduced to just two parameters, and a Bayesian predictive approach is used to construct the multivariate control chart.

There are many semiconductor manufacturing processes that may induce a covariance structure that conforms to the parsimonious covariance that we are proposing (cor-

relation that degrades with distance in space). Some examples of these are an LPCVD (low pressure chemical vapor deposition) reactor that is used for the deposition of polysilicon or nitride, or a PECVD (plasma enhanced chemical vapor deposition, Engle 1980) reactor for the deposition of nitride or oxide. To further illustrate this, we consider a passivation process for semiconductor manufacturing, where each of k variables (measures of deposition) represents a value at a specific location in a passivation tube, and because of the interaction between the plasma and the reactant gases flowing down the tube, the correlation among the variables might decay with distance between these locations. Moreover, taking equal variability at the locations might further reduce the number of parameters. Diffusion processes, which would include oxidation and annealing to drive dopants into silicon substrates, constitute another step that may be explained by our method. Finally, single wafer processes may be another application for our method; wafer etching typically produces radial patterns on the wafer that induce some form of spatial correlation across the wafer.

Hotelling (1947) pioneered the work on multivariate control charts. In particular, the values plotted on the control charts are usually statistics related to his well-known T^2 statistic. His work has led to several groundbreaking applications of multivariate methods to industrial problems. Indeed, there are many examples of multivariate processes where the use of separate individual charts would not have detected out-of-control conditions; see, for example, Ryan (1989) and Montgomery (1991). If there is prior information about the form of the covariance matrix, the performance of these multivariate charts can be improved substantially.

Multivariate control charting is an area of much research activity. Sullivan and Woodall (1996) compared a number of classical multivariate control charts for individual observations, and showed that the procedure of pooling all data to estimate the mean vector and the covariance matrix is not effective in detecting a shift in the mean vector because the covariance matrix is badly estimated. Thus, they considered several alternatives to estimate the covariance matrix. Our method differs from theirs in that, with hindsight, we can estimate the covariance matrix efficiently if there is parsimony. Nonparametric approaches are described by Liu (1995) in which unlike in standard approaches (e.g., Tracy et al. 1992) both location and scale shifts are detected. Also Liu and Tang (1996) describe control charting using bootstrap methods. However, our data can be modeled using a multivariate normal distribution, and we desire to incorporate prior information about the covariance matrix. In addition, the special case of multivariate statistical process control with

several identical process streams (i.e., commensurate measurements for the same multivariate observation) is similar in spirit to our applications (see, for example, Runger et al. 1996).

Alt (1982) described two distinct phases for constructing control charts. The first phase, screening, is retrospective, and in this phase, control limits are constructed with historical data from a stable process. It is assumed that the principles of rational subgroups are used to minimize the effects of assignable causes within subgroups and to maximize such effects between subgroups. The second phase, monitoring, is prospective, and in this phase, the control limits, which were obtained from historical data, are used to detect departures from the process standards as future subgroups are taken. The correlation structure of the data should be taken into consideration in both of these phases.

Tracy, Young, and Mason (1992), henceforth referred to as TYM, provide a simple description of a standard method for constructing multivariate control chart. Let $y_1, y_2, \dots, y_n \sim N(\mu, \Sigma)$, where y_i is the vector of k measurements at the i^{th} run, μ is the $k \times 1$ mean vector and Σ is the $k \times k$ covariance matrix of the process. Letting $\bar{y} = \sum_{i=1}^n y_i / n$ be the sample mean vector and $S = \sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})' / (n-1)$ be the sample covariance matrix, for screening with $n > k+1$, the statistic $Q_i = (y_i - \bar{y})' S^{-1} (y_i - \bar{y})$ is compared with the control limits derived from $nQ_i / (n-1)^2 \sim \text{Beta}(k/2, (n-k-1)/2)$. For monitoring a future multivariate observation y_f the charting statistic $T^2 = (y_f - \bar{y})' S^{-1} (y_f - \bar{y})$ is compared with the control limits derived from $n(n-k)T^2 / k(n+1)(n-1) \sim F(k, n-k)$. TMY (1992) used the .01, .50 and .99 quantiles of the beta distribution and the F -distribution to perform screening and monitoring respectively.

In contrast our method, which uses a predictive approach, is based on a simple idea. Given sample data the predictive distribution of the next observation is computed, and it is used to derive the control limits when the process is in control using a charting statistic analogous to Hotellings' T^2 statistic. When further observations are taken from the process and if the observed charting statistic falls outside the control limits, the process is said to be out of control. Furthermore, in principle, normality is not required to execute this method. However, like TYM we study a multivariate normal process.

For both methods the condition $n > k$ can be a nuisance. For example, for a typical application with $k=5$, there are $k(k+3)/2 = 20$ parameters to be estimated, and if n is not much larger than k , the parameters μ and even worse Σ cannot be estimated efficiently. See Boyles (1996) for an extensive discussion of this problem. If the application involves only commensurate variables, one can hope to reduce the number of parameters in the covariance matrix by utilizing the physical properties of the process.

Unfortunately, with the parsimonious covariance matrix the distributions of Q_i and T^2 are not as simple as stated by TYM. Thus, while this parsimony in reducing the number of parameters may lead to improved precision, there is some loss in simplicity. Fortunately, the gain in precision will outweigh the loss in simplicity. Boyles (1996) demonstrated increased precision in the parsimonious covariance structure which occurs in applications with a multidimensional lattice (a grid of points in a coordinate system). He also pointed out that there is no exact distribution for setting control limits and proposed an approximation to the distribution of Hotelling's T^2 charting statistic. We provide essentially an exact distribution.

Using a sampling based method, it is conceptually simple to obtain the distribution of the charting statistic under the parsimonious covariance structure, together with any useful prior information that may be obtained from historical data. The charts are constructed using the Metropolis-Hastings algorithm (Chib and Greenberg 1995) through a statistic analogous to Hotelling's T^2 . During the screening phase the control limits are constructed using a cross-validation and future monitoring is done using the analogous Hotelling's T^2 statistic with the control limits established from the cross-validation.

Thus, we describe a Bayesian multivariate chart appropriate for the passivation process in the semiconductor industry. The rest of the paper is organized as follows. In the next section we describe the semiconductor data and a preliminary data analysis. In the following section we describe the methodology; in particular, the Bayesian model and the new charting statistic are discussed. We then describe control charting in our application on the semiconductor data. We also describe a simulation study to assess the performance of our method and to compare it with the TYM method. The last section contains conclusions. In this paper, we omit the technical details which can be obtained from Nandram and Ramirez (2005).

Semiconductor Data

The computer industry has made substantial progress in technology and capability in the past twenty years; complex computations that used to be accomplished by a computer that filled an entire room can now be done using a laptop computer. One of the primary reasons the computer industry has advanced so fast is due to the technological advances in silicon based integrated circuits. An integrated circuit is smaller than a dime, and can contain millions of transistors with metal lines thinner than a human hair. These metal lines carry high density current. It is very difficult to manufacture integrated circuits; a typical process consists of hundreds of steps. Passivation is one such step and has motivated this work. The passivation process is performed to deposit an insulating, protective layer over the entire device after metalization. The passivation layer can be deposited onto the wafers in a furnace (tube) that uses plasma-enhanced chemical vapor deposition. A graphite boat in the tube, which was used for our application, is configured to hold 96 4-inch wafers.

The passivation tube is partitioned into zones in both the width and the length of the tube (see Figure 1). There are 12 rows that run down the length of the passivation tube with 5 pairs of rows that run back-to-back. Each row can hold 8 wafers and each wafer is placed with its back against the slat in an upright position. Nitrous oxide, silane, and diluent are pumped into the tube where the wafers are loaded, gas flows parallel to the wafers' surface, and finally a plasma is ignited. The plasma energy breaks down the reactant gases, allowing chemical vapor deposition to occur on the wafer surface, thus forming the passivation layer. Monitor or bare silicon wafers are strategically placed at three locations in one of the middle rows near the gas source, center, and door in the tube.

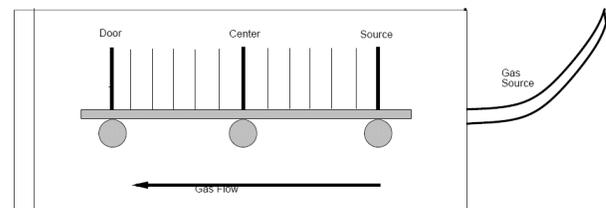


Figure 1. Location of the wafers in the passivation tube with gas pumped in

At the end of each run, five thickness measurements are taken from the top, center, left, right, and bottom positions of each of the three wafers. The quality of the process is assessed by studying the variability of these film thickness measurements. A *NanoSpec*®, which relies

on the interference of light, is used to measure the thickness of the film. The equipment is capable of measuring film thicknesses from 100 angstroms to 50,000 angstroms (1 angstrom is 1×10^{-8} centimeters). The Nanospec is a very reliable instrument for measuring commercially prime (monitor) wafers. However, there may be some difficulties in taking measurements if extreme variation in surface conditions of the underlying substrate exists. Also, the interaction between the plasma and gas flow down the tube, along with temperature gradients, causes variation in film properties, such as film thickness.

The average of the five measurements on each wafer is usually studied. However, the three wafer averages that represent the three tube locations are usually highly correlated. This is expected because all the wafers in the tube are simultaneously exposed to the same conditions for each run. Therefore, it is sensible to consider the three measurements as a three dimensional vector. It is believed that, for all practical purposes, the variability at the three locations can be taken equal, and that the correlation among the locations decay with distance between the locations. However, the thicknesses of the coating at the three locations are expected to be different.

The process is monitored by taking a number of runs in time. The objective is to use the thickness data to construct the control limits for the process and then to monitor the future of the process using the established control limits under the prescribed covariance structure.

Our data set, collected by a US based semiconductor manufacturer, consists of 26 runs from a passivation tube where each run consists of the average film thicknesses on 3 monitor wafers located at the source, center and door of the passivation tube (i.e., $k=3$ and $n=26$). For convenience, the original measurements were divided by 1000 for our analysis. We are unable to present the original data here because of confidentiality issues. However, in Table 1 we have provided a masked data set which can be used for exploration.

Thus, we present a preliminary analysis of the 26 three-dimensional measurements. The sample pairwise correlation between the wafer averages at the source and center is 0.89, between the center and the door is 0.87, and between the source and the door is 0.74, indicating that the population correlation matrix has the requisite structure. The sample variance at the source, center, and door are 0.64, 0.41, and 0.38, respectively; the average estimated variance is 0.48. These estimated variances indicate that the corresponding population variances can be taken equal.

Table 1: Masked data at the source, center and door of the passivation tube

| Source | Center | Door |
|--------|--------|------|
| 2.05 | 2.07 | 2.07 |
| 2.10 | 2.10 | 2.07 |
| 2.05 | 2.05 | 1.98 |
| 2.07 | 2.03 | 1.92 |
| 2.05 | 2.04 | 1.89 |
| 2.06 | 2.07 | 2.09 |
| 2.08 | 2.05 | 2.04 |
| 2.02 | 2.00 | 1.72 |
| 2.10 | 2.09 | 2.12 |
| 2.57 | 2.52 | 2.40 |
| 2.39 | 2.20 | 2.09 |
| 2.17 | 2.17 | 2.12 |
| 2.31 | 2.34 | 2.30 |
| 2.22 | 2.30 | 2.28 |
| 2.09 | 2.08 | 2.00 |
| 2.29 | 2.39 | 2.36 |
| 2.23 | 2.32 | 2.28 |
| 2.21 | 2.28 | 2.22 |
| 2.15 | 2.12 | 2.18 |
| 2.13 | 2.21 | 2.21 |
| 2.07 | 2.12 | 1.87 |
| 2.16 | 2.14 | 2.15 |
| 2.10 | 2.07 | 2.02 |
| 2.09 | 2.09 | 1.99 |
| 2.15 | 2.17 | 2.14 |
| 2.18 | 2.24 | 2.21 |

Note: These data give answers different from the original data.

In addition, we perform a preliminary analysis to determine differences between the locations in the tube and to get rough estimates of σ^2 and ρ using a repeated measures model (i.e., the correlation is the same between any two locations). Letting y_{ij} denote the components of y_i , we used PROC MIXED in SAS to fit the model

$$y_{ij} = \theta + \kappa_i + \nu_j + e_{ij},$$

where $\kappa_1, \kappa_2, \dots, \kappa_n \sim N(0, \delta^2)$ and independently $e_{11}, e_{12}, \dots, e_{nk} \sim N(0, \sigma^2)$ with fixed effects $\theta + \nu_j$ at location $j, j=1,2,3$. Then, the means of y_{ij} are estimated by 10.78, 10.82 and 10.53 for the pump, center and the door respectively. The estimate of the common variance of y_{ij} is 0.478 at the source, center and door and the estimate of the correlation is 0.816. Also, a test of $\nu_1 = \nu_2 = \nu_3$ gives a p-value of .001, indicating that the thicknesses at the three locations are significantly different.

We also fitted the full covariance matrix (i.e., the TYM model) using the method based on the quantiles of a beta distribution described by Rencher (1995, Sec. 4.4.2). The quantile plot shows six points departing from normality in which runs 8, 10, 11 and 21 are included. Both tests for skewness and kurtosis reject normality with all 26 observations. When we repeated the test with 22 runs, about three points in the quantile plot departed from linearity, and the test for kurtosis rejected normality. Next, we present our methodology and apply it to these data.

Bayesian Methodology

In this section, we show how to construct a Bayesian multivariate control chart. First, we describe a model appropriate for commensurate measurements with a parsimonious covariance structure. Second, we use these samples to construct a Mahalanobis distance analogous to Hotelling's T^2 which, in turn, is used to obtain the control chart (i.e., the lower control limit, center line and upper control limit). Screening is done using a Bayesian cross-validation. Again, note that the technical details are presented in Nandram and Ramirez (2005).

Modeling

Quite generally we consider k measurements taken at a particular time point. (The k measurements constitute a k -dimensional vector.) We also consider n historical runs to construct the control charts. We incorporate prior information about the parsimony of the covariance matrix using a Bayesian model.

A multivariate normal distribution is not an unreasonable candidate distribution for our measurements. Thus, starting with multivariate normality for the measurements, we take iid

$$y_1, y_2, \dots, y_n \mid \mu, \Sigma \sim N(\mu, \Sigma), \quad (1)$$

where $\mu = (\mu_1, \mu_2, \dots, \mu_k)'$ is the mean vector and $\Sigma = \sigma^2 \Gamma$ is the covariance matrix with Γ equal to the correlation matrix (assuming equal variability at each position in the tube).

With appropriate prior information, one can choose the correlation matrix to have a reduced number of parameters. An application of particular interest to us is the one in which $\Gamma = (\rho_{jj'})$ and

$$\rho_{jj'} = \rho^{|j-j'|}, \quad j, j' = 1, 2, \dots, k, \quad 0 < \rho < 1;$$

this suggests that locations further away have smaller correlations.

Although our model is specific, our method could be applied in the same spirit to other correlation structures. For example, a very simple covariance structure arises in a equicorrelated and equivariance situation where $\Sigma = \sigma^2(\rho I + (1-\rho)J)$ with I the $k \times k$ identity matrix and J the $k \times k$ matrix of ones. A more complicated situation can arise if the jj' element of Σ is $\sigma_j \sigma_{j'} \rho^{|j-j'|}$, $j, j' = 1, 2, \dots, k, 0 < \rho < 1$. Here the variances are different and correlation falls off with distance, and therefore, our specific model is a special case of this more general model. But even in this more general situation there could be significant parsimony with even a moderate number of variables.

In any application the assumption of multivariate normality, together with our proposed covariance structure, must be assessed. Note that the results of Tracy et al. (1992) do not apply to this model. In particular, if the parsimonious covariance matrix is used, the sample covariance matrix must also have this same pattern and the distributions used for charting are not as described in the review of the first section of this paper.

For priors, we take $p(\mu) = 1$, $p(\sigma^2) \propto \sigma^{-2}$ and $p(\rho) = 1, 0 < \rho < 1$; in this case the performance should be similar to the one based on maximum likelihood estimation if one can carry out an accurate analysis.

Under the Bayesian model specified by (1) and the prior distributions just defined we will construct a statistic analogous to Hotelling's T^2 for control charting. The posterior distribution of this statistic is estimated using a sampling based method through the Metropolis-Hastings sampler (see Nandram and Ramirez 2005), and the control limits are obtained using the appropriate percentiles of the charting statistic.

The Charting Statistic

Let $\Omega = (\mu, \sigma^2, \rho)$ be the set of all parameters and $y = (y_1, \dots, y_n)$ be the set of all data. Let y_f be the vector of the k measurements of a future run. Then our charting statistic is

$$D_f = (y_f - E(y_f \mid y))' (cov(y_f \mid y))^{-1} (y_f - E(y_f \mid y)), \quad (2)$$

where $E(y_f \mid y)$ and $cov(y_f \mid y)$ are the posterior mean and covariance of y_f . Given Ω we assume that $y_f \sim (\mu, \Sigma)$ and independent of the past runs y . Then, it is easy to show that $E(y_f \mid y) = E(\mu \mid y)$,

$$cov(y_f \mid y) = E(\Sigma \mid y) + cov(\mu \mid y). \quad (3)$$

More importantly, the posterior predictive distribution of y_f given y is

$$\pi(y_f | y) = \int_{\Omega} f(y_f | \Omega) \pi(\Omega | y) d\Omega. \quad (4)$$

Note that the posterior distribution of $\Omega | y$ is needed to compute the posterior distribution of D_f in (2). Because the posterior distribution of $\Omega | y$ is intractable, the posterior distribution of D_f is also intractable.

Under the model it follows from Bayes' theorem that the joint posterior distribution $\pi(\Omega | y)$ is given by

$$\pi(\Omega | y) = \pi_0(\Omega) \ell(\Omega | y) \quad (5)$$

where $\pi_0(\Omega)$ is the prior distribution given by $\pi_0(\Omega) \propto \sigma^{-2}$ and $\ell(\Omega | y)$ is the likelihood function

$$\ell(\Omega | y) \propto (\sigma^2)^{-nk/2} (1 - \rho^2)^{-n(k-1)/2} \times \exp\{-(2\sigma^2(1 - \rho^2))^{-1} (a - 2\rho b + \rho^2 c)\}$$

with

$$a = \sum_{i=1}^n \sum_{j=1}^k (y_{ij} - \mu_j)^2, \quad b = \sum_{i=1}^n \sum_{j=1}^{k-1} (y_{ij} - \mu_j)(y_{i,j+1} - \mu_{j+1}) \quad \text{and} \\ c = \sum_{i=1}^n \sum_{j=2}^{k-1} (y_{ij} - \mu_j)^2. \quad (\text{See Appendix A of Nandram and Ramirez 2005 where the likelihood function is derived.})$$

Thus, the posterior distribution $\pi(\Omega | y)$ does not exist in closed form since the proportionality constant (i.e., the integral of (5) over the $(k+2)$ -dimensional space) cannot be obtained analytically, and can be difficult to find using numerical integration. Samples from the posterior distribution in (5) are required to construct the posterior distribution of the charting statistic.

Using the composition method (e.g., Tanner 1993) in (4) we obtain an estimate of the posterior distribution of y_f in two steps (a) samples of the posterior distribution of Ω are drawn and (b) for each Ω drawn, a y_f is drawn from $N(\mu, \Sigma)$. While step (b) is obtained in a straightforward manner, step (a) is much more difficult. However, to obtain D_f in (2) we must first obtain $E(y_f | y)$ and $cov(y_f | y)$ in (3) which are obtained in an obvious manner from the samples drawn in (b). Step (b) is accomplished side-by-side with step (a), but, of course, it is not a part of the sampling process in which the posterior distribution of Ω is obtained; see Nandram and Ramirez (2005). By steps (a) and (b), we get samples from the posterior distribution of D_f in (2).

Finally, for control charting we obtain the $100\alpha^{th}$ percentile, 50^{th} percentile and $100(1-\alpha)^{th}$ percentile of the D_f values. (Tracy et al. 1992 took $\alpha = 0.01$.) Data screening for outliers is performed by a Bayesian cross-validation; see Nandram and Ramirez (2005) for details.

Henceforth, for convenience we will call this new method the NR method, and we will denote the charting statistic D_f by T_{NR} .

We screen the data of n runs for outliers by performing a Bayesian cross-validation. The i^{th} run is simply deleted, and it is treated as a future value. Then, the D_i as in (2) are computed for all runs, deleting each in turn. Values of D_i that are much different from the others are suspect and should be removed from the data to establish the control limits. This technique is analogous to the one described by Gelfand et al. (1992) to obtain diagnostics for Bayesian models; see Nandram and Ramirez (2005).

To monitor the manufacturing process in the future we simply calculate a single D_f for each future run which becomes known and compare this value of D_f with the already established control limits.

Data Analysis and the Semiconductor Data

Now, we apply our methodology to the semiconductor data. Specifically, we show how to perform multivariate control charting, and we describe a simulation study to assess the performance of our method.

Analysis of the Semiconductor Data

First, we perform a brief goodness of fit procedure. Using standardized deleted residuals, we first assess the fit of the Bayesian model with the parsimonious covariance using cross-validation as described by Gelfand et al. (1992). Let $y_{(i)}$ denote the vector of all runs when the i^{th} (y_i) run is deleted, $v = E(y_i | y_{(i)})$ the posterior mean, $B = cov(y_i | y_{(i)})$ the posterior covariance of $y_i | y_{(i)}$, and $B^{-1} = C'C$ Cholesky's decomposition of B^{-1} with C an upper triangular matrix. Then defining $z_i = C(y_i - v)$, the standardized residuals z_{ij} are approximately independent and normally distributed a posteriori. Figure 2(a) shows the normal probability plot of the standardized residuals for all runs. There are five points which deviate from the 45° line. Two of the largest residuals are associated with run 8 with residual values -2.84 and 2.88, and the other three largest residuals are associated with runs 10, 11 and 21. (Run 8 has the smallest measurement at the door.)

When we omitted all four runs associated with these five residuals, we obtained a much improved normal probability plot (Figure 2(b)). (Note that there are 78 points in Figure 1(a) and 66 points in Figure 2(b).)

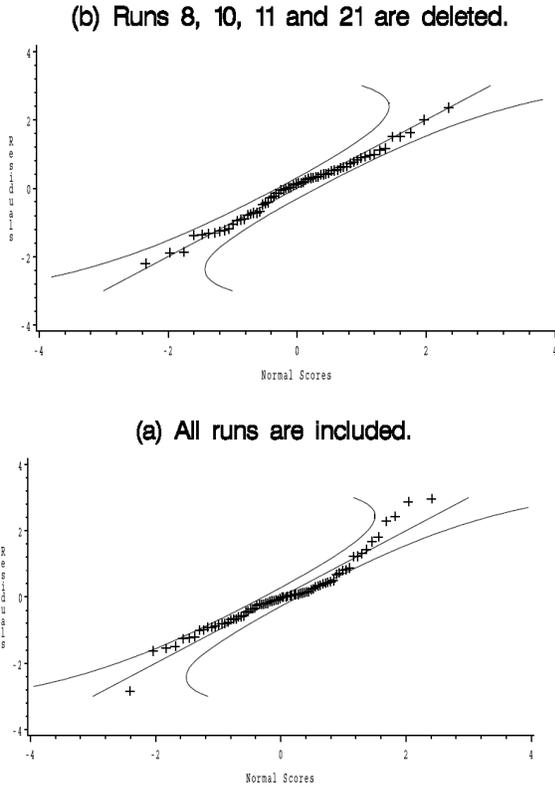


Figure 2. Normal probability plot of standardized residuals including the expected 45 degree line and the 95% pointwise critical bands

Figure 3 shows the control charts based on T_{TYM} and T_{NR} . Notice that the out-of-control runs are different for the two charting statistics. Of course, the limits based on T_{TYM} are larger than those based on T_{NR} . Figure 4 shows the control chart with the one out-of-control run indicated by T_{NR} in Figure 3 removed. Now there are three out-of-control runs which are indicated by both charting statistics. It is comforting that the two charts for T_{TYM} and T_{NR} look similar with the limits for T_{TYM} being larger.

On removing runs 8, 10, 11 and 21 Figure 5 indicates that the process is in control. The process engineer indicated that these outliers were probably due to poor electrical contact of the wafers to the boat in the tube. Thus, the control limits are the established limits which will be used for future monitoring. For T_{TYM} the control limits are 0.12, 2.71, 16.6 and for T_{NR} the control limits are

0.10, 2.21, 12.5. The posterior means of T_{NR} and T_{TYM} are 2.85 and 3.85 respectively. Not surprisingly the posterior standard deviations of T_{NR} and T_{TYM} are 2.44 and 3.57 respectively, an increase of 46% of T_{TYM} over T_{NR} .

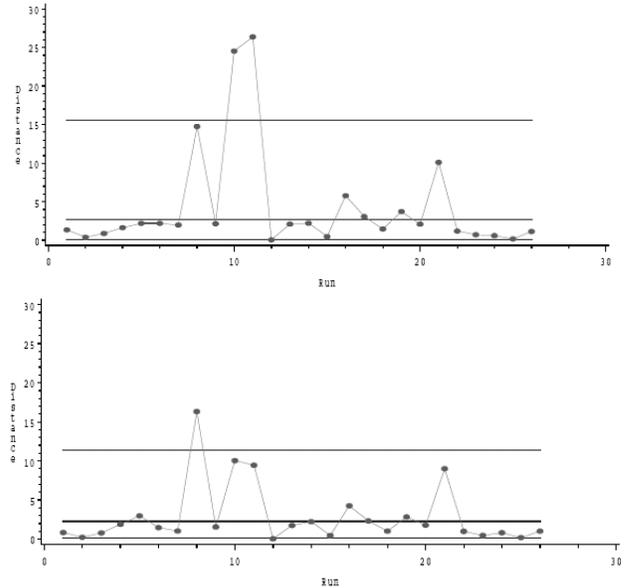


Figure 3. Control charts for the NR method (bottom panel) versus the TYM method (top panel)

Finally, in Figure 6 in pretense we consider the 26 runs as future runs with the established control limits. This is the same as Figure 3 with the limits shown in Figure 4. Thus, the out-of-control runs are different for the two methods.

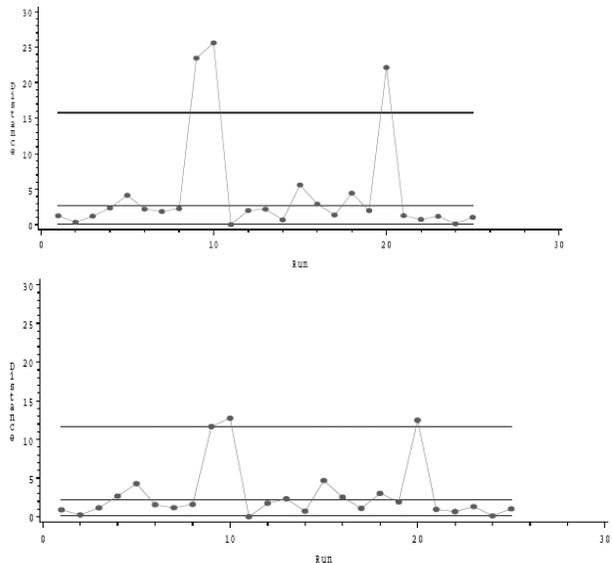


Figure 4. Control charts for the NR method (bottom panel) versus the TYM method (top panel) without run 8

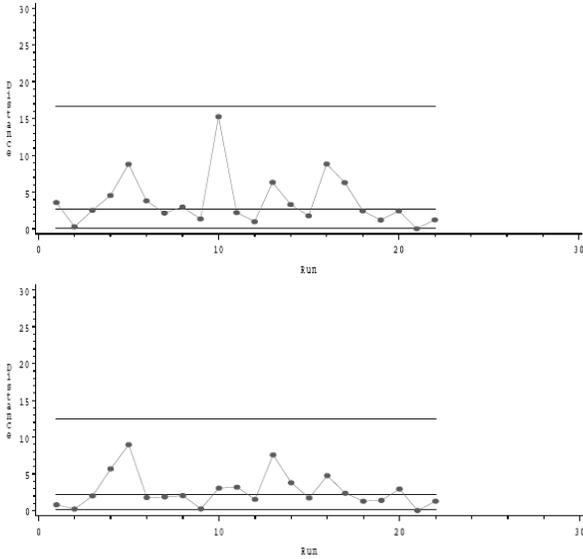


Figure 5. Control charts for the NR method (bottom panel) versus the TYM method (top panel) without runs 8, 10, 11 and 21

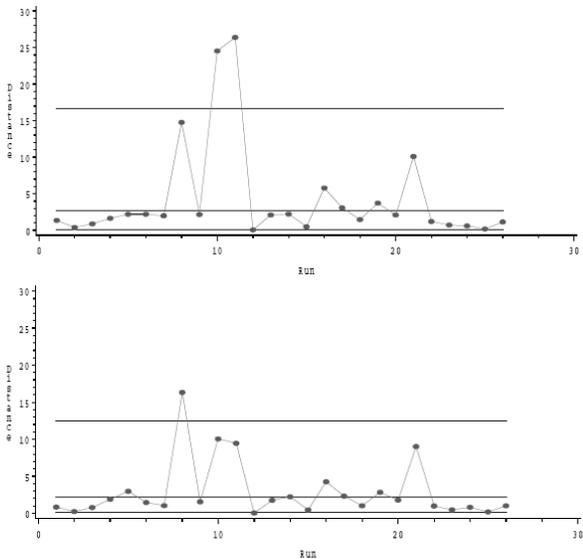


Figure 6. Control charts for the NR method (bottom panel) versus the TYM method (top panel) for the 26 runs after establishing control (i.e., without runs 8, 10, 11 and 21)

A Simulation Study: Detection of Out-of-control Runs

In this section we use simulated examples to study how well the NR method can detect out-of-control values as compared with TYM. In the simulated examples the components of μ are all set at zero for convenience, and the covariance matrix has specified forms (see below). Then, we generated a random sample of $n = 26$ k -variate normal random variables with the specified mean vector and covariance matrix. (There are $n = 26$ runs in

our application.) In all examples we used noninformative priors. We compare the average run length for the process under the NR model and the TYM model when there is an out-of-control process shift. In our simulated examples we took $\sigma^2 = 1$, $\rho = .80$, $\mu_1 = \mu_2 = \dots = \mu_k = 0$ for $k = 2, 3, \dots, 8$. Note, in particular that the data are generated under the parsimonious model.

Let LCL and UCL denote the $100(1-\alpha)\%$ control limits for either the NR method or the TYM method. We consider a known shift of δ_i for the i^{th} component of the mean vector for the stable process, and we represent the multivariate vector of shifts by $\delta = (\delta_1, \dots, \delta_k)$. Then, for a shifted observation, denoted by y_f , $y_f - \delta$ follows the original model

$$y_f - \delta \mid \mu, \Sigma \sim N(\mu, \Sigma), \tag{6}$$

where Σ refers to the full covariance matrix for TYM or the reduced covariance matrix for NC. Then, the posterior distribution of $y_f - \delta$ is

$$p(y_f - \delta \mid y) = E(p(y_f - \delta \mid \mu, \Sigma)) \tag{7}$$

where the expectation is with respect to the posterior distribution of μ and Σ given y obtained from the Metropolis-Hastings algorithm.

We computed the average run length (ARL) when there is a shift of δ . Let N be the number of runs required for the process to go out of control (i.e., for the process to go outside the interval (LCL, UCL)). Then the ARL is given by

$$ARL = E(E(N \mid y, \Omega)) = E(E(N \mid \Omega)),$$

where $\Omega = (\mu, \sigma^2, \rho)$. But given Ω , N has a geometric distribution with probability $p(\Omega)$ depending on Ω with expectation $p(\Omega)^{-1}$. Thus,

$$ARL = E(p(\Omega)^{-1} \mid y). \tag{8}$$

We obtain a Rao-Blackwellized estimator of the ARL as follows: for each iterate $\Omega^{(m)}$, we draw 5,000 values y_1, \dots, y_{5000} under the shifted distribution and count the proportion out of control, denoted by $p(\Omega^{(m)})$ for the m^{th} iterate. Then the Rao-Blackwellized ARL is given by

$$M^{-1} \sum_{m=1}^M p(\Omega^{(m)})^{-1},$$

where $M=2,000$ is the number of iterates from the Metropolis-Hastings algorithm. Increasing the sample size to 10,000 showed virtually no difference in the point estimates. We use a similar method for TYM.

Others have used a Markov chain approximation to determine the ARL of a multivariate statistical process control chart; see, for example, Runger and Prabhu (1996). Our method for computing the ARL needs essentially no approximation.

We consider two cases. In the first case, we perturbed all k variables in the same way and calculated the probability of detecting the stated shift. We study shifts of δ_i (σ is set at 1) of 0.0 to 3.5 by steps of 0.25 for 2 to 8 variables. In the second case, we perturbed only the first of the k variables.

For the false alarm rate (i.e., the out-of-control probability for an in-control process) with reasonably small error, we obtain the .02 nominal value for both methods. When all variables are perturbed at 0 shift (i.e., in control), the ARLs range from 49-66 for the NR method and 57-59 for the TYM method. There is no pattern as the number of variables increases from $k=2$ to $k=8$. Except for $k=4$ when $ARL=66$ and $k=7$ when $ARL=49$, for all values of k the ARLs for NR and TYM are very similar. The NR ARLs are very similar when all components are perturbed to when only the first component is perturbed, but the TYM ARLs are slightly larger when only the first component is perturbed, ranging from 61-64. These are approximately in concordance with the .02 nominal value.

In Table 2 we present the results for shifts of 0.5, 1.0, 1.25, 1.5, 2.0, 2.5 sigma. It is not surprising that for $k=2$ the two methods are very similar for both cases. In general the ARLs for TYM are much larger than those

for NR. The exceptions are for $k=5$ and case 2 at shifts 1.25 and beyond but the differences are not alarming. At the same value of k for case 1 at shift 1.25 compare 29 and 50 and at shift 1.5 compare 20 and 44. In many examples the ARLs for NR are less than half those for TYM. For $k=6, 7$ or 8 and for both cases the NR method is substantially better than the TYM. For example for $k=8$ and any case the ARL for the NR method is less than a quarter of the ARL for the TYM method. Observe that when only the first variable is shifted, the ARLs are smaller. The NR method is expected to perform better than the TYM method as the number of variables increases because the number of parameters for TYM method increases dramatically but increases only by one for each additional variable for the NR method.

Conclusion

In this paper, our main contribution is in screening and monitoring multivariate semiconductor data in which it is likely that there is parsimony in the covariance. In particular, we have introduced a Bayesian cross-validation method into statistical process control, and we have shown that it is feasible to reduce the number of parameters in a k -dimensional covariance matrix to just two parameters.

This structure with a reduced covariance matrix is suitable not only for statistical process control for semiconductor manufacturing, but it may also be appropriate for data structures arising in areas such as agriculture, geostatistics and image analysis.

Table 2: Comparisons of ARLs by number of variables, case, shift and method

| k | Case | Shift | | | | | | | | | | | |
|---|------|-------|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|
| | | 0.5 | | 1.0 | | 1.25 | | 1.5 | | 2.0 | | 2.5 | |
| | | NR | TYM | NR | TYM | NR | TYM | NR | TYM | NR | TYM | NR | TYM |
| 2 | 1 | 50 | 53 | 38 | 38 | 28 | 27 | 19 | 18 | 8 | 8 | 4 | 4 |
| | 2 | 47 | 47 | 20 | 18 | 10 | 9 | 5 | 5 | 2 | 2 | 1 | 1 |
| 3 | 1 | 52 | 56 | 34 | 40 | 23 | 27 | 14 | 17 | 6 | 6 | 3 | 3 |
| | 2 | 48 | 55 | 21 | 33 | 11 | 20 | 6 | 11 | 2 | 4 | 1 | 2 |
| 4 | 1 | 56 | 57 | 32 | 42 | 21 | 28 | 13 | 16 | 5 | 6 | 3 | 3 |
| | 2 | 54 | 57 | 25 | 33 | 15 | 17 | 8 | 9 | 3 | 3 | 2 | 2 |
| 5 | 1 | 52 | 59 | 38 | 58 | 29 | 50 | 20 | 44 | 9 | 21 | 4 | 9 |
| | 2 | 48 | 59 | 25 | 26 | 15 | 11 | 9 | 6 | 3 | 2 | 2 | 1 |
| 6 | 1 | 56 | 61 | 40 | 60 | 28 | 52 | 18 | 40 | 7 | 16 | 3 | 6 |
| | 2 | 55 | 62 | 32 | 58 | 19 | 46 | 11 | 30 | 4 | 11 | 2 | 4 |
| 7 | 1 | 42 | 62 | 24 | 64 | 16 | 57 | 10 | 43 | 4 | 16 | 2 | 6 |
| | 2 | 40 | 61 | 20 | 66 | 12 | 65 | 7 | 56 | 3 | 28 | 2 | 12 |
| 8 | 1 | 53 | 63 | 38 | 71 | 27 | 71 | 17 | 68 | 7 | 38 | 3 | 17 |
| | 2 | 51 | 64 | 32 | 72 | 20 | 68 | 12 | 61 | 4 | 27 | 2 | 10 |

Although the TYM method is the simplest among the classical methods, our example illustrates that one cannot apply the TYM method naively. In fact, both the TYM and the NR methods need careful model checking. Our model fits better in our example. The NR method is superior in detecting all shifts as compared to the TYM method. While the difference is smaller for shifts of 0.5 sigma or so, this difference is substantial for larger shifts and larger number of variables.

We have shown that when the NR model holds, there is substantial improvement over the TYM method. In particular, our method is able to detect out-of-control values faster than the TYM method. This is true for all shifts and any number of variables, more so for larger shifts and larger number of variables. Specifically, this is beneficial for the passivation process in the semiconductor industry.

Acknowledgments

The research was funded by UNITRODE Corporation, Merrimack, New Hampshire. We are also indebted to Edmund Burke, a Process Engineer at UNITRODE who generously contributed the data and insight about the passivation process.

REFERENCES

- Alt, F. B. 1982. Multivariate Quality Control: State of the Art, in *ASCQ Quality Congress Transactions*, 886-893.
- Boyles, R. A. 1996. Multivariate Process Analysis With Lattice Data, *Technometrics*, 38: 37-49.
- Chib, S. and E. Greenberg. 1995. Understanding the Metropolis-Hastings Algorithm, *The American Statistician*, 49: 327-335.
- Engle, G. M. 1980. Plasma Enhanced Chemical Vapor Processing of Semiconductor Wafers, *US Patent # 4, 223, 048 September 16, 1980*.
- Gelfand, A. E., D.K. Dey, and H. Chang. 1992. Model Determination Using Predictive Distributions and Implementation via Sample-based Methods, *Bayesian Statistics 4*, eds. J.M. Bernardo, A.P. David, and A.F.M. Smith, Oxford University Press, New York, NY, 147-167.
- Hotelling, H. 1947. Multivariate Quality Control, *Techniques of Statistical Analysis*, eds. C. Eisenhart, M. Hasty, and W. A. Wallis, McGraw-Hill, New York, NY, 111-184.
- Liu, R. Y. 1995. Control Charts for Multivariate Processes, *Journal of the American Statistical Association*, 90:1380-1387.
- Liu, R. Y. and J. Tang. 1996. Control Charts for Dependent and Independent Measurements Based on Bootstrap Methods, *Journal of the American Statistical Association*, 91: 1694-1700.
- Montgomery, D. C. 1991. *Introduction to Statistical Quality Control* (2nd ed.), Wiley, New York, NY.
- Nandram, B. and B. Ramirez. 2005. A Bayesian Predictive Inference for Multivariate Control Charting, *Technical Report*, Department of Mathematical Sciences, Worcester Polytechnic Institute.
- Rencher, A. C. 1995. *Methods of Multivariate Analysis*, Wiley, New York, NY.
- Runger, G. C., F.B. Alt, and D.C. Montgomery. 1996. Controlling Multiple Stream Processes with Principal Components, *International Journal of Production Research*, 34:2991-2999.
- Runger, G. C. and S.S. Prabhu. 1996. A Markov Chain Model for Multivariate Exponential Weighted Moving Averages Control Chart, *Journal of the American Statistical Association*, 91: 1701-1706.
- Ryan, T. P. 1989. *Statistical Methods for Quality Improvement*, Wiley, New York, NY.
- Sullivan, J. H. and W.H. Woodall. 1996. A Comparison of Multivariate Control Charts for Individual Observations, *Journal of Quality Technology*, 28: 398-408.
- Tanner, M. A. 1993. *Tools for Statistical Inference*, Springer-Verlag, New York, NY.
- Tracy, N. D., J.C. Young, and R.L. Mason. 1992. Multivariate Control Charts for Individual Observations, *Journal of Quality Technology*, 24: 88-95.