Statistical Feedback Control of a Plasma Etch Process

Purnendu K. Mozumder and Gabriel G. Barna

***Abstract*—This paper presents the methodology developed for the automatic feedback control of a silicon nitride plasma etch process. The methodology provides an augmented level of control for semiconductor manufacturing processes, to the level that the operator inputs the required process *quality* characteristics (e.g. etch rate and uniformity values) instead of the desired process conditions (e.g., specific RF power, pressure, gas flows). The optimal equipment settings are determined from previously generated process/equipment models. The control algorithm is driven by the in-situ measurements, using in-line sensors monitoring each wafer. The sensor data is subjected to Statistical Quality Control (SQC) to determine if deviations from the required process observable values can be attributed to *noise* in the system or are due to a sustained *anomalous* behavior of the equipment. Once a change in equipment behavior is detected, the process/equipment models are adjusted to match the new state of the equipment. The updated models are used to run subsequent wafers until a new SQC failure is observed. The algorithms developed have been implemented and tested, and are currently being used to control the etching of wafers under standard manufacturing conditions.**

I. Introduction

SEMICONDUCTOR processing is typically performed with machine-dependent, static process menus. These menus are usually generated by some process characterization techniques (e.g. Taguchi, Response Surface Modeling [1]–[3]) and the optimum process then becomes the specification for the manufacturing operation. The resulting process observables are then typically tracked by an SQC technique. When this system detects an out-of-control situation, the process is re-centered via engineering intervention or the hardware is cleaned up and recommissioned in hopefully the original in-spec state.

This mode of operation can track process drifts, but only reacts to them once an SQC limit for one of the process observables has been violated. Since these observables are generally measured after-the-fact (i.e., with a significant delay in time after the process), a large number of wafers can potentially be misprocessed before the out-of-specification processing condition is recognized and corrected. However, recent advances in the field of plasma processing have allowed significant improvements to be made toward the rapid run by run, as well as real-time control of such processes [4]–[6]. This has been partly enabled by the development of in-situ sensors for the monitoring of process observables (e.g., etch rate via ellipsometer measurements [7], line-width change by scatterometry [8]). There has also been a trend toward modeling the process observables (e.g. etch rate, selectivity, uniformity) as a function of the process parameters (e.g., RF power, pressure, gas flows), by means of empirical models [9], [10]. The merging of these technologies has paved the way for the model-based control of plasma processes. For the controllers to be effective, the lag between the occurrence of a fault and a statistically confident detection of the fault must be minimized. A quick detection scheme reduces the amount of time for which the process is out-of-control, which minimizes the number of potential out-of-specification product wafers. However, the SQC technique has to have a low out-of-control false alarm rate, since an SQC violation will lead to a correction step. Finally, the controller needs to be protected against instability in the presence of unexpected and large deviations in the equipment state.

Such a robust SPC procedure was one of the specific goals of the MMST (Microelectronics Manufacturing Science and Technology) program currently being carried out at Texas Instruments. The intent of this portion of the program was to modify the requirements of the manufacturing operation to utilize an equipment-independent process specification; where this specification contains the required results from the process, not the standard machine settings. In order to run a process in this mode, with the ability to modify the process parameters between wafers to keep the observables in-spec; one needs the following basic components:

1) sensors for the measurement of the process results,
2) an in-line SQC technique for detecting when these observables have gone out-of-specification, with a statistical confidence,
3) a model that relates the process observables to the process parameters, and
4) an optimizer to calculate the new settings for re-centering an out-of-specification process.

This paper will describe the details of these four basic components, and the application of this new scheme to the plasma etching of a nitride film. For ease of presentation, these four basic components will be described in a different order than presented above. The next section (Section II) comprises the description of the nitride etch process, the equipment configuration, and the modeling experiments used to create the control models. The use of the sensor to deduce the responses of interest will also be briefly described in Section II. The following section (Section III) contains the strategy and implementation used to control the nitride etch process. Finally, the behavior the controller under several fault scenarios is presented in Section IV.

II. Equipment Models and Sensors

This paper will be focused on the etching of a nitride film in a PBL (Polysilicon Buffered LOCOS) stack (2400 Å silicon

Manuscript received October 22, 1992; revised June 25, 1993. This work was supported in part by the Air Force Wright Laboratory and the DARPA Microelectronics Technology Office under Contract F33615-88-C-5448.

The authors are with Semiconductor Process and Design Center, Texas Instruments Inc., Dallas, TX 75243.

IEEE Log Number 9214294.

0894-6507/94$04.00 © 1994 IEEE

nitride, 500 Å polysilicon and 90 Å silicon dioxide). The requirements for this etch are a high nitride etch rate and low etch nonuniformity, and a small line width loss. Since the dominant layer in the PBL stack is the nitride, almost all of the line loss can be attributed to the nitride etch. It is also important to ensure that the during the nitride etch no part of the wafer has the polysilicon completely etched, so as to expose the underlying oxide, since the oxide is readily etched by the $CHF_3/CF_4/O_2$ etch chemistry used for the nitride etch. This would cause a lateral encroachment of the subsequent field oxidation on the active area of the transistor. There are two possible methods of avoiding the etching of the poly: a very high etch selectivity of nitride to poly, or a very uniform etch. The former is not achievable (the selectivity is at most 4 : 1), and therefore having a low nonuniformity is an important target for the etch.

As stated previously, some of the key requirements for the rapid run by run feedback control of processes are the in-situ sensors for detecting process shifts and the models required to re-center the process after such a shift has been observed. The following experiments were designed and run for the PBL etch on a TI-designed and built Advanced Vacuum Processor (AVP) [11], to generate the empirical models to be used in the model-based process control operation of this reactor. In addition, data had to be obtained to evaluate the correlation between the slope of the endpoint curve at the nitride endpoint and the uniformity of the etch. The reason for this was to see if this slope of the endpoint curve could be used as a sensor for the etch rate uniformity.

The primary equipment controls for the AVP are the chamber pressure ($Pressure$), delivered RF power (RF), four gas flows (CHF_3, CF_4, O_2, and Ar), the helium chuck pressure ($He - chuck$), and the rector substrate temperature ($Substrate$). The output parameters of interest for the PBL etch are the etch rate (ER), across wafer etch rate nonuniformity (NU), and line width change (LW). There are four linewidth structures on the wafers where linewidth change can be measured. Although all four sites were modeled only one of the structures was used by the controller. Details of these structures and the rationale for the choice of a particular linewidth model for control are presented in Sections II and III. A description of the observables, and the corresponding sensor interpretation, is presented in Section II; only the ER and NU could be observed using the in-situ sensors present in the AVP.

Experimental Setup

The AVP is a single-wafer plasma reactor with the following attributes:

1) Process chamber enclosed by quartz walls,
2) RIE configuration, with a wide electrode gap,
3) 6″ wafer on a cooled electrode with a He-chuck, clamped face-down to the top electrode by means of three quartz pins, and
4) Gas inlet through a quartz tube, with a centrally-located and clamped "table-top" gas distribution baffle over the end of the tube.

The hardware is under full computer control, with a 386 PC with UNIX. All mass flow controllers (MFC's) are calibrated by the internal 10-point calibration method, with the set-points automatically defined from this calibration curve. Prior to running the experiments, the reactor hardware was stabilized and characterized. Software was implemented for the closed-loop control of all control parameters, and the standard PBL etch process was determined to be reproducible. This is to emphasize that stable hardware and reproducible process conditions are pre-requisites for the model-based process control mode of operation described in this paper.

Responses and Measurement Methodology

The only on-line sensor available on this particular AVP is a single wavelength monochrometer, used to derive a real-time *optical emission signal* (OES) during the etch. With the monochrometer tuned to a spectral line that corresponds to the emission of a chemical species that is depleted or generated during the etch, the resulting intensity versus time curve is known as the endpoint trace (EPT). Of the two process observables required for the process control strategy described in this paper, ER (etch rate) is readily obtained from the standard interpretation of the EPT. NU (across wafer etch rate nonuniformity), on the other hand, requires that a correlation be made between the EPT and the nonuniformity. If this correlation is sufficiently high then the OES will be an effective measure (low variance predictor) of both etch rate as well as nonuniformity; a desirable set of conditions for multivariable control.

The monochrometer was tuned to the CN^* emission line at $388.3nm$. An example trace is shown in Fig. 1. At this wavelength the changing plasma chemistry produces a clear end-point (the large decrease in intensity at $\sim 65s$ in the Fig. 1), which then triggers a timed overetch and subsequently the RF shutdown. The etch rate is readily calculated from the automatically determined endpoint time and the known thickness of the nitride film.

Based on previous experience and theoretical considerations [12], it was anticipated that the slope of the EPT, at the nitride etch end-point, would correlate to the nonuniformity of the etch. A set of experiments was performed specifically to obtain the correlation between the endpoiñt slope and the nonuniformity of the etch. Since it was determined from earlier experiments that the nonuniformity was sensitive to only three of the input parameters ($Pressure$, RF and CHF_3), a full factorial experimental design was run, appended with three replicate center points ($2^3 + 3 = 11$ runs). Two wafers were run for each experimental point; the first was etched past the end-point with a datalogger used to capture the EPT at a $0.2s$ sampling interval. The second wafer was only etched to 75% completion and was subsequently analyzed with 81 point pre- and post-etch film thickness measurements to obtain a value of the spatial etch uniformity across the wafer. Although a simple measurement of the slope could have been used, the electronically logged data lent itself to the sophisticated analysis of the slope. Using the methods of smoothing [13], differentiation, and feature extraction, metrics

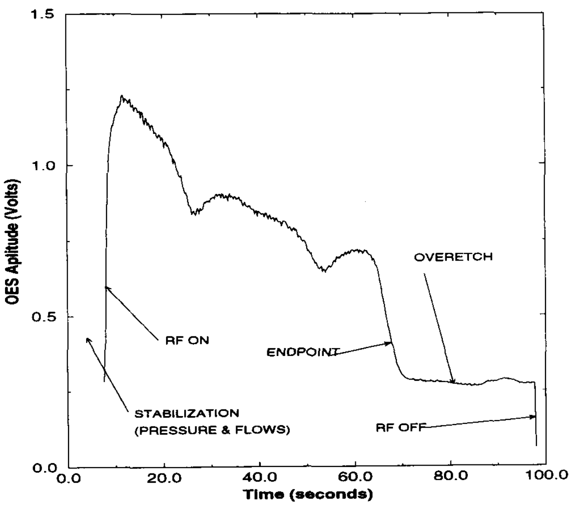


Fig. 1. Optical emission signal trace for $388.3nm$ line during PBL etch. (All equipment controls are held constant during the etch. RF is shut down at 98 s.)

of slope at the end-point were determined and then correlated to the $(max. - min.)$ and $standard\ deviation$ (σ/μ) metrics of nonuniformity. A correlation coefficient value of 0.8 was observed between predicted, in terms of the slopes of the EPT (at the end-point), and the experimentally observed nonuniformity (σ/μ).

Although linewidth (LW) control is significant for this etch, the *in situ* critical dimension (CD) sensor developed in the MMST program is not available on this AVP. This implies that all line width measurements have to be routed to off-line metrology tools for measurements. Since this entails very time consuming pre- and post-measurements[1], which prohibits the utilization of this metric for rapid run by run or real-time process control, the LW models have therefore been taken out of the control loop. However, to assure that the optimization of the ER and NU does not violate LW considerations, the off-line LW models are included as constraints in the process optimization calculations. Since the linewidths were not observables only one of LW models (the one with the best model fit — Section II) is used in the controller.

Equipment Modeling

With *in-situ* real-time sensors providing data on the process observables, the next requirement for statistical process control is the availability of process models. The models are necessary to determine the optimum starting point for the process, and to vector/drive the process observables back toward the target values as the hardware drifts with time. Preliminary experimental studies illustrated that polynomial, quadratic models had sufficiently high goodness of fit for the outputs of etch rate, nonuniformity and the line width loss for a limited range of variation in the input values[2].

[1] The linewidth of the resist before and after the etch are determined using a top down scanning electron microscopy (SEM) image.

[2] Two line losses and two space gains were modeled. For the experiment, however, only one of the line loss models was used.

TABLE I
INPUT PARAMETER RANGES FOR PBL MODEL.

Parameter	Units	Low	Center	High
$Pressure$	mT	100	150	200
RF	W	300	400	500
CHF_3	$sccm$	30	40	50
CF_4	$sccm$	40	60	80
O_2	$sccm$	10	15	20
Ar	$sccm$	-	100	-
$He-chuck$	T	-	2.5	-
$Substrate$	$^{\circ}C$	-	0	-

The range of the control parameters chosen for this modeling experiment was determined to be the range over which the inputs will be allowed to vary when the process is subjected to SQC. Statistically designed experiments were generated via the commercially available experimental design package, ECHIP^{TM} [14]. Table I shows the input ranges for the five control parameters which were varied and the values for the three which were fixed (and thus are not a part of the model). The "standard" process was kept at the center of the hyperbox.

In order to generate the necessary data for the modeling of the ER, NU and line width loss (LW), the experimental procedure was as follows. The experiments generated by ECHIP^{TM} (a 31 wafer *D-optimal* [15] experimental design, with 26 unique experiments, along with the 5 replicates) were actually run in three separate sets, each with slightly different wafers which were optimized for the measurements that had to be made.

Set #1:	
Substrate	Silicon
1000Å	Oxide
2000Å	Poly
2400Å	Nitride
No Resist	

These wafers were run past the nitride etch endpoint in order to determine the endpoint time under the individual process conditions.

Set #2:	
Substrate	Silicon
1000Å	Oxide
2400Å	Nitride
No Resist	

These wafers were run to about 75% of the previously determined nitride endpoint time. These wafers were evaluated for etch rate nonuniformity by an 81 point measurement of nitride thickness on a *Prometrix SpectraMap*.

Set #3:	
Substrate	Silicon
1000Å	Oxide
2000Å	Poly
2400Å	Nitride
Moat pattern	CRB Positive Resist

TABLE II
MODEL STATISTICS FOR PBL MODELS.

Parameter	ER	NU	LW-1x1	LW-3x3	LW-1x5	LW-3x5
R^2	0.951	0.927	0.947	0.901	0.894	0.905
Residual Std. Dev.	4.095	8.682	0.0106	0.0092	0.0181	0.0162

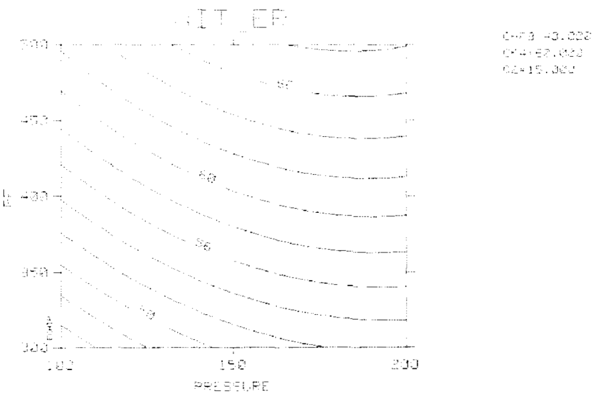


Fig. 2. ER versus RF and pressure.

These wafers were etched to just past the nitride endpoint. Line width data was obtained from pre/post SEM measurements on 5 dies per wafer, with 5 points on each die.

Based on the analysis of the data by ECHIP^{TM}, it was determined that of the six observables, five were modeled well (i.e. the variance of the replicates accounted for most of the variance in the residuals). The residual standard deviation and the R^2 for the modeled parameters are presented in Table II. The NU has poor predictive capability, due to the residual standard deviation of approximately 8.6%. The LW-1x5 model had the worst fit amongst the LW models. LW-1x1 ($1.2\mu m$ line) had the best model fit and was used for the controller. (The NxM notation after the last two parameters refers to the die/position of the measurement. Die #1 is close to the wafer flat, #3 is in the middle of the wafer, and #5 is at the top of the wafer. Although several sets of linewidth data were measured, only four were modeled.)

The contour plot representation of the etch rate is shown in Fig. 2. The contour is plotted in the coordinate system that had the greatest effect on the specific observable. The etch rate shows essentially a pure RF power dependence at higher pressure, with more of a RF/PR interaction at lower pressure. The coefficients of all the terms in the quadratic model of each observable were extracted and incorporated into the model-based control algorithms of the process control software.

III. CONTROL STRATEGY

From the viewpoint of the operator, there is a significant difference between running a process under machine control or under process control. In the former, one specifies the hardware control parameters; in the latter, the process observables are specified. In order to execute the process under process control, the strategy shown in Fig. 3 has been implemented. The specific steps in the execution of this strategy are as follows:

1) Query the user for a set of targets on ER (etch rate), NU (nonuniformity) and LW (line width).
2) Determine the optimal settings based on the RSM's. This is the "target-to-settings" step. The process of determining the optimal equipment settings based on the process/equipment models, and the required/desired values of the observables (and nonmeasured product parameters) is termed as target-to-settings. The objective function used is the sum of squares of the difference between the predicted and target values. The optimization is carried out using an encapsulated version of the optimizer NPSOL [16].
3) Measure the ER and NU by observing the sensor (optical emission signal from the monochrometer) data or off-line measurements. In this case, the off-line SEM measurements were not made. The software is set up, however, to take into account the SEM measurement if the user is able to measure them between successive runs.
4) Perform SQC based on the (observed—model predicted) output for ER and NU. This is based on the use of two charts; the *Moving Average* and the *Moving Standard Deviation* with a sample size of 4.
5) Based on the results of the SQC test (described in detail in the next section) the subsequent options are as follows:
   a. If the output is outside spec limits, stop processing and begin diagnosis;
   b. else, if the process violates control limits, determine whether ER or NU has failed, adapt the constant term in the corresponding model, and continue processing with the new control values;
   c. else, continue processing with the present model.

To avoid getting trapped in a local optimum during the target-to-setting optimization, 20 starting points generated by a *Latin Hypercube Design* [17] are used. The best optima from the 20 runs are chosen. Although this does not guarantee a global optimum, it minimizes the probability of being stuck in a local optimum. The number 20 was chosen as a compromise between time to find the settings and the probability of hitting the global optimum.

SQC Charts

In order to analyze the process data obtained from the in-situ measurement of the ER and NU, it is necessary to analyze the trend in the mean and standard deviation over several runs. This led to the use of the Moving Average ($\overline{X}$) and

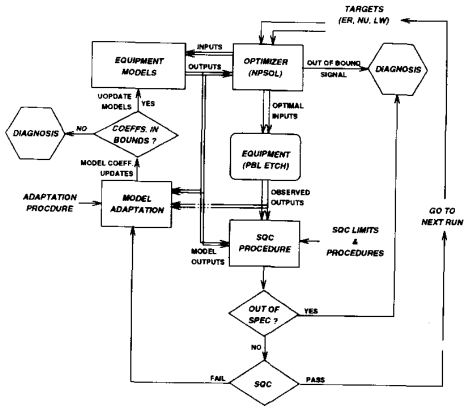


Fig. 3. Schematic of SPC strategy for PBL nitride etch on AVP-PBL.

Moving Standard Deviation (s) SQC methods[3]. The mean and variance for n consecutive runs are plotted on an $\overline{X}$ and s charts, respectively. The $\overline{X_k}$ and s_k after the kth run, using n samples, are calculated as:

$$\overline{X_k} = \frac{1}{n}\sum_{i=1}^{n} X_{k-i+1} \tag{1}$$

$$s_k = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} X_{k-i+1}^2 - n(\overline{X})^2\right]} \tag{2}$$

The control limits and the specification limits have to be generated for the SQC chart. The SQC limits depend on the process' intrinsic variability, or the expected behavior of the process under noise. The specification limits are set up by the process and control engineers as the difference between the observed and predicted that signals a large malfunction and cannot be treated as a small shift or drift in the process/equipment. The specification limits are generated using prior history and simulations. The values serve to determine the limits on the SQC chart and to target expected values. Based on the mean and standard deviation, 4 limits are set up: Upper Action limit (UAL), Upper Warning Limit (UWL), Lower Warning Limit (LWL), and Lower Action limit (LAL). Using these set of limits, Western Electric Co. (WECO) [18] rules are applied to the data to determine when an "out-of-control", or "out-of-specification", signal is to be generated by the SQC system.

Since the SQC method was to be generalized for all processes, and different processes inherently possess different dynamics and noise levels the choice of the sample size, n in (1) and (2) is fairly critical. It can be used to tradeoff the control stability with the delay in "out of control" signal. A small sample size allows the quick detection of changes in the equipment state, but has an increased probability of false alarms. A larger value of n, on the other hand, usually means lesser chance of false alarms and hence less jitter in the controller. However, it also introduces a smoothing effect that causes an increased lag between the time when an abnormal shift is observed versus when the SQC signals an "out of control" situation and the model tuning event is triggered. A given fixed sample size will not be the optimal choice in terms of the lag time to detect a fault and false alarm rate for different processes/equipment. Therefore, the ability to vary sample size was included in the SQC method. To facilitate the use of variable sample size, a method of generating the limits for the variable sample size has been developed. The effect of sample size upon the statistical coefficients used in setting the SQC limits is given in Appendix V. Since this PBL etch process on the AVP has been found to be reasonably stable over time, i.e., the machine and the process parameters do not vary significantly over a short time, the sample size used for averaging over time was chosen as 4. The coefficients corresponding to a sample size of 4 are also given in Appendix V. In order to calculate the SQC limits an estimate of the mean and standard deviation are needed. The following section explains the methodology for choosing the values for $\overline{X}$ and s.

[3] The Individual (X) charts were not used for SQC to minimize the probability of false alarms.

It was possible to consider multiple SQC charts with multiple sample sizes (possibly one with a small and another with a large sample size), to attempt to reduce both the probability of false alarms and detect shifts or drifts with minimal lag. However, this approach was not taken since it would require a more complex algorithm to determine out of control conditions, and would probably need human intervention whenever the results from the two control charts were in conflict. Since part of the requirements were to keep the system reasonably free of human intervention, and simple for the operator to understand, we have limited ourselves to a single SQC chart.

Model-Based SQC and Model Adaptation

Model-based SQC has to accommodate the users need for requesting a different output for every wafer, if needed. This in turn means that the successive output values cannot be presumed to be samples from a known distribution, and hence SQC cannot be performed on the outputs. However, the model prediction is expected to track the observed output value for different targets, and hence the difference between the model prediction and the output values is a good indicator of shift in the process.

The basis of the SQC procedure lies in the regression equation that is used to estimate the models' coefficients.

$$\hat{y} = f(\mathbf{x}) \tag{3}$$

$$y = \hat{y} + \epsilon \tag{4}$$

where

1) y is the actual output from the AVP (say ER),
2) $\hat{y}$ is the output predicted by the corresponding model,
3) $f(\mathbf{x})$ is the equation of the model (the RSM for ER),
4) $\mathbf{x}$ is the vector of input parameters, and
5) ϵ is the error term, that signifies the noise in the prediction error.

The term ϵ is the key parameter for performing model based SQC. The theory of regression states that ϵ is the *confounded* measure of the pure experiment error and the error in the models (lack of fit) [19]. It is a random variable distributed *normally* with $N(0, \sigma)$ assuming insignificant lack of fit[4] [15], where σ is the standard deviation of the error variable. This in turn means, as per (4), that the difference in the predicted and observed outputs$(y - \hat{y})$ should a sample from the distribution for ϵ. This forms the basis of the SQC charts, with mean value of 0.0 and standard deviation of σ [20], [21]. An unbiased estimate of σ^2 is the *residual mean square error*, generated by the ANOVA decomposition of the experiment design matrix [19].

In this work, a univariate SQC scheme is used where each observed output variable (ER and NU) is treated independently for SQC failures. A single model parameter is updated for each of the corresponding models. Optimal process conditions in terms of the input variables are generated to simultaneously maintain all the three outputs to target. The control strategy can be summarized as a multivariable control technique with a single coefficient (constant per model) update feedback policy, using an univariate SQC scheme.

The Individuals (X) chart[5] is more susceptible to false alarms than the $\overline{X}$ charts. Since the etch process requires a minimal probability of false errors, hence SQC based on the X charts was not implemented. However, the user is provided the X chart via the graphical interface, a part of the user interface for running SPC on the nitride etch AVP. A brief description of the implementation of the SPC algorithms is provided in Section IV. It is also understood that using the $\overline{X}$ instead of the X charts for SQC leads to a smaller probability of false alarms at the cost of introducing a delay in detecting a fault that causes a shift in the equipment state [22].

If there is an SQC failure for the $\overline{X}$ chart, it is an indication that the process has shifted and that the models need adaptation. Since a univariate algorithm is used, only a single parameter can be adapted based on the present and past values of the difference between the predicted and observed output. This parameter can either be a level shift (the constant term in the equation), or a gain term (a single coefficient). For simplicity, we have implemented the shift in the constant term. On an SQC failure for the Moving Average, the constant for model corresponding to the output parameter causing the SQC violation is updated. However, the magnitude of this update is bounded, so that the models do not chase a run-away process. Once the model is adapted, target-to-settings is achieved using the new model.

The $\overline{X}$ chart trends follows the X chart with a delay. To prevent the controller from oscillating, it is required that the effect of the model tuning be backed out of the past samples. If this were not done the mean value would show a large deviation due to past samples, when in fact the model has adjusted for the deviation and brought the process to target. A similar effect would be noticed in the Moving Standard Deviation chart where a tuning would be marked by a large "glitch" in the SQC chart. Both these effects would cause false SQC failures and make the controller oscillate to instability.

[4] All the models for the process had insignificant lack of fit. Most of the residual variance could be attributed to pure error terms, estimated from the replicate runs.

[5] The X chart is one where the deviation of the observed values from predicted are plotted for each run.

Although both past and present data is used to calibrate/adapt the new models, forgetting factors are used to weight the present data more than the past. A filtered value of the mean over the last 25 samples is used to estimate the change in the model's constant term. To remove the effect of the lag between the step where the machine state changed and the step where the SQC was triggered, "exponentially-weighted backing out" scheme was developed. The tuned parameter in the model is backed out past previous model adaptations/SQC failures, up to 25 wafers; and the resulting scheme serves two purposes:

- It prevents the data from the wafers which fall in the lag between the abnormal event and the trigger to cause an spurious SQC following the model adaptation, and
- It minimizes the over- or under- biasing of the calculated model update during frequent SQC failures.

The choice of the forgetting factor is important to the stability of the SPC scheme. The methodology for choosing an appropriate factor is not well defined in the literature [23], [24]. Prior knowledge of the process stability and behavior is the key to the choice of its value. Since the forgetting factor is used to exponentially weight the past, its value remains $0 \leq \lambda \leq 1$. We have used a filter factor of 0.55 (or forgetting factor of 0.45), based on simulated experiments.

After a wafer is processed any one, or multiple, output parameters may fail the SQC for $\overline{X}$. The models corresponding to the responses that fail SQC must be adapted. For the remainder of the parameters, the models are assumed to be unchanged. This is done to simplify the strategy, knowing full well that one of the parameters for which there was no failure may be critically close to failure and may fail in the next few steps.

The theory of regression states that if the model has a good fit then the lack of fit is small compared to the pure experiment error (noise in the system), and the prediction error standard deviation remains constant over the *convex hull* of the experiment space [25]. This implies that whenever a SQC alarm is generated based upon an s chart failure, the process models represented in (3) and (4) are probably no longer valid, and new models may have to be created. The source and the corrections for a s chart violation are not as simple to analyze [26] as the $\overline{X}$ chart violations. Therefore no automated action is built into the controller.

It is known from literature that the use of univariate techniques can affect the false alarm and failure to detect out of control signal probabilities when the outputs under SQC are correlated [27]–[30]. It is known from the models that the controlled outputs (ER and NU) for the PBL process are correlated. However, we have chosen a univariate SQC technique for simplicity. It was also noticed during routine operation of the controller that in spite of the univariate SQC charts, the controller behaved stably and was able to keep the process under control specifications. For the subsequent controllers developed under MMST, multivariate SQC schemes have been used.

Implementation

The sensor interpretation and SQC/SPC algorithms described above have been implemented as a part of a program called *AVPSPC*. *AVPSPC* is a generic model based statistical process control software package driven by user defined files. This software was written, in its entirety, in *C*. The platform is *Unix* with *X Window System* and *Motif*. The target machine is a *Micronics 386 PC* with *SCO Unix*. The main development system is a *SUN SPARCstation*.

C code has been implemented that takes the datalogger values[6] of the EPT (endpoint trace) and calculates a value for nonuniformity (as a percentage of the etch rate). The code has been integrated with the machine control software on equipment. The OES (optical emission signal) becomes a sensor for the in-situ measurement of the etch rate and etch rate uniformity for each wafer, providing values of these metrics at the conclusion of the etching of each wafer.

The off-line activities that precede SPC are entered using a menu- and form-driven interface that allows the control engineer to input the controller configuration (i.e., SQC procedure, sample size and tuning mechanism), forgetting factor, SQC limits, spec. limits, etc.. The details of the software and configuration procedures will not be described in detail in this paper.

IV. Experiments and Results

With the appropriate sensors in place, the static RSM modeling completed, the control strategy defined and all the software tools integrated, the last test prior to the implementation of the process control mode of operation was the verification of the system's ability to correctly recover from deliberately induced hardware errors.

SPC Verification Experiments

A set of fifty PBL etch experiments were run on the same AVP in order to test the controller concept and the associated hardware and software. In order to expedite the violation of the SQC rules, which is the requirement for initiating the retuning of the model, some of the hardware parameters included in the modeling were deliberately misadjusted in a fashion that was unknown to the controller, in several stages.

1) Wafer Hardware status prior to wafer run.
2) W-01 Normal.
3) W-10 RF Limit changed such that delivered RF was 20% greater than that specified by the menu.
4) W-13 RF Limit further adjusted such that delivered RF was 40% greater than that specified by the menu .
5) W-16 O_2 Calibration Table also misadjusted so that O_2 flow was 60% of the menu value.
6) W-17 O_2 Calibration Table again misadjusted so that O_2 flow was 50% of the menu value.
7) W-18 O_2 Calibration Table again misadjusted so that O_2 flow was 200% of the menu value.
8) W-26 RF Limit table and MFC calibration values reset to correct values.

Results

The results of the verification experiments mentioned in the previous section are presented in the form of SQC charts that illustrate the (actual-predicted) values for each wafer. For each of the parameters ER and NU, three SQC charts are shown: Individuals (X), Moving Average ($\overline{X}$), and Moving Standard Deviation (s). Each chart shows the cumulative run number on the abscissa[7], and the individual values or statistical moments of the deviations between the model prediction and actual values of the outputs[8], on the ordinate. Plotted in the Individual, Moving Average and the Moving Standard Deviation charts are the individual values, the 4 sample means and the 4 sample standard deviations, respectively. The six WECO SQC limits and the specification limits (outside the 3σ limits) [18] are also specified on each SQC chart . Figs. 4 and 6 are the Individual charts for the ER and NU. These charts illustrate the behavior of the output caused by equipment state changes which were unknown to the controller, and the results of the subsequent action by the controller to re-center the process such that the output comes closer to the predicted value. The Moving Average and Moving Standard Deviation charts are useful for analyzing the controller's behavior (e.g., the lags, the reason for SQC violations, etc.).

The following is a detailed analysis of the SQC charts pertaining to ER, as ER is far more sensitive to the misadjustments than NU, hence causing significant SQC failures and corresponding control actions. During all these runs, the target values of ER and an NU were $50\text{Å}/s$ and 5.0%, respectively. It is important to note that the generalized SQC procedure does not require that the targets remain constant, the SQC is performed on the deviation from the model prediction. If the optimizer is able to generate the equipment control settings that results in the model prediction being same as the target then the SQC is performed on deviation from target, and hence does not depend on the target value chosen. These target values have been modified after the verification experiments were completed.

The first 5 (up to Run#10) wafers were run to baseline the process and start the Moving Average (Figs. 5 and 7) and Moving Standard Deviation (Figs. 8 and 9) charts that require a 4 sample delay at startup. It is important to note that model tuning is based solely on the SQC performed on the Moving Average. Therefore even though the individual values of the deviation were large, the mean took 4 samples to realize the large deviation. A closer observation of the control charts reveal three features:

- The mean lags the individual samples since there is an averaging over 4 samples. The averaging thus results in a delay in detecting a deviation and prompting an SQC violation.
- The smaller the deviation from the model prediction, larger the time lag before the WECO rules triggers. This

[6] The control parameters and observables are sampled at regular intervals in time and stored in a database.

[7] Note that although the starting run numbers for the different charts are slightly different due to the graphical interface software, the run numbers are consistent between charts.

[8] The mean and standard deviation charts start up with a lag since it takes a 4 sample start up before a valid value of the mean or standard deviation can be calculated.

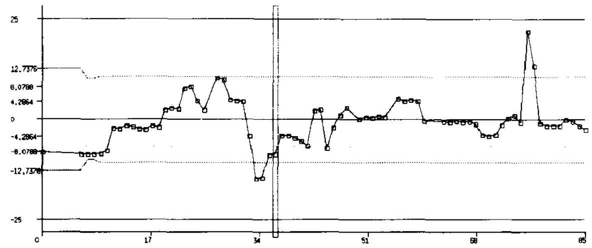


Fig. 4. Individuals chart for (actual—model) of etch rate.

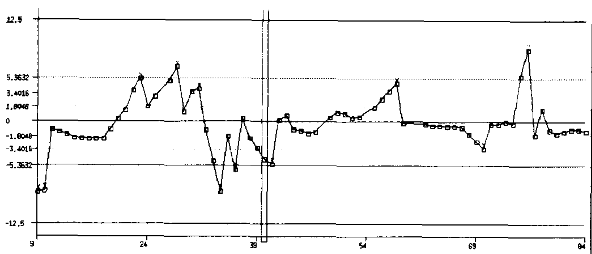


Fig. 5. Moving average chart for (actual - model) of Etch Rate.

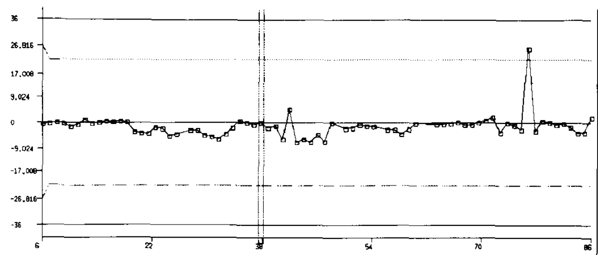


Fig. 6. Individuals chart for (actual - model) of nonuniformity.

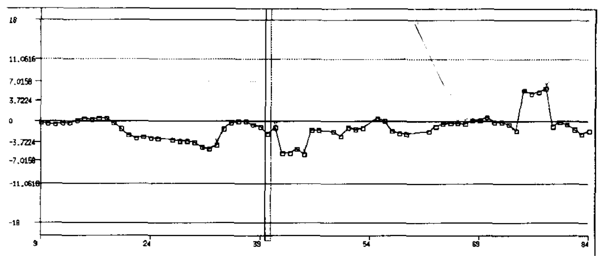


Fig. 7. Moving average chart for (actual - model) of nonuniformity.

is because the WECO rules require a certain probability of rejecting the null hypothesis (that there is no abnormality). The number of samples required to attain the probability is lower for larger deviations.

- The larger the deviation, the smaller is the lag between the mean and the individuals.

Once the Moving Average chart is able to detect the deviation (Run#10), it immediately signals an SQC failure in the chart (by placing an "x" in the data point) and invokes the model tuning procedures. The effect of the model tuning is seen as the first large discontinuity (before Run#11) in the chart in Fig. 4, where tuning the model makes the deviation from target for the next wafer close to 0. The deviation from the model prediction is approximately $-2.0Å/s$. Since this is within the σ of the model (which is set to 3.5Å), there is no further SQC violation seen, as the SQC procedures interpret the small deviation as noise rather than an abnormality[9].

The next significant discontinuity (Run#20) was seen for wafer W-10 where misajdusting the RF limits in the calibration table caused the RF generator to increase delivered RF power, which resulted in an etch rate increase, but only enough to place it just above the upper 1σ band. This would require 4 out of 5 samples between 1 and 2 σ for SQC to fail. There would be a 5 to 7 sample delay before the failure was to be registered on the Moving Average chart. Therefore the RF table was further misadjusted (Run#23), causing the generator to deliver higher power, and subsequently causing the deviation in ER to increase further (W-13). This caused a SQC violation with a 2 sample delay, causing the model to retune and bringing the ER close to target (Runs#25 & #26). Notice, that the Moving Average follows the SQC chart with a delay, as explained in Section II. To prevent the controller from oscillating, the effect of the model tuning is backed out of the past samples. This backing out of the model in turn meant that the mean was suddenly close to target after a model tuning event is triggered, in a manner uncharacteristic of the gradual nature of Moving Average charts. Much smaller glitches, characteristic of an exponentially weighted filtering, are encountered on the control charts.

The next two wafers were run with the MFC miscalibrated (along with the RF). At this point, the optimizer requested a flow of $10sccm$ but only about $5sccm$ was delivered. This low flow rate could not be controlled to within the hardware control specification limits, causing the hardware controller to signal flow problems, leading to a machine error. Therefore, these runs are not recorded in the SQC charts. The SPC system, however, journals these machine failures. On wafer W-18 (before Run#27) the flow meter was uncalibrated in the other direction, so as to provide greater than the requested flow. The effect was similar to the RF misadjustment - the controller was able to recover from the problem (as observed on Run#29).

Finally, both the RF and O_2 MFC were reset back to their original, correctly calibrated values (W-26), at which point the observed ER was much smaller than the corresponding prediction (after Run#31). SQC violation was observed after 2 wafers (when the mean value picked up the deviation - Run#34). Of the 4 runs used to calculate the Moving Average, 2 of them had small deviations. Consequently, the Moving Average did not "catch up" with the value of the deviation (i.e. did not attain the maximum value). However, the deviation was large enough to trigger an SQC alarm based on the value of the Moving Average and the controller was activated. As a result the final value of the required model adjustment calculated by the SPC feedback loop (with a filter factor of 0.55, and using only 2 observations with the abnormal deviations of observed values from model prediction) was

[9] The 8 on same side of mean rule, which takes care of such small biases in the data, was accidentally shut of by repeated Moving Standard Deviation failures in the NU chart. This problem was later debugged and worked correctly for the NU at a later point in the experiment.

smaller than that of the value required to bring ER to target. This is evident in the charts, where the Individuals chart shows a large deviation in spite of the tuning (Runs#35 & #36), and the Moving Average chart shows "spikes" that are due to the Moving Mean gradually attempting to catch up with the deviation of the Individuals. Two more tunings were necessary to bring the outputs back close to target (before Runs#37 & #42).

With this portion of the process control verification experiments successfully completed, approximately 10 more wafers were run with differing target values, ER ranging from $60\text{Å}/s$ to $45\text{Å}/s$, and NU from 20% to 5%. In most of the cases deviations of the actual values from the model predictions were small. However, it is observed that the deviation from model prediction were not always close to zero (especially when a high ER and a high NU were jointly desired). This may have resulted from an inherent limitation of the model tuning methodology. Where as in reality the change in the equipment state may have required more than one of the coefficients to have changed, the present strategy only allows for the model to be "shifted", or "translated". Therefore, it is conceivable where a slope change (e.g. coefficient for RF or O_2) was actually required, a change in the constant term would suffice if the change was small, but not if the change was large. This limited tuning policy would eventually show up in bad model predictions especially at the edge of the model's domain.

The SQC charts for the NU (Figs. 6 and 7) show that NU is less sensitive to the perturbations than ER. However, a downward trend is noticed due to the RF change (after Run#19) (faster ER resulting in smaller NU), which was then tuned back to almost zero deviation from the target value (Run#35). A similar change was observed when back sides of wafers were etched (Runs#43 – #47), which was then corrected for by the SPC system by tuning the model (Run#48). It is also important to notice, from the Moving Standard Deviation chart (Fig. 9), that the NU was in fact better controlled than what the model predicted. Most of the moving standard deviation points in the chart are smaller than 4.0%, and a fairly large number is below 3.2%, whereas the residual standard deviation for the corresponding model is 8.6%. This may have been due to the fact that the equipment state has drifted to a regime where the NU is fairly insensitive to small perturbations. However, this caused SQC failure on the Moving Standard Deviation charts where the standard deviation frequently went below the lower control limit. Normally, such an error is indicative of the model not being valid. These failures were "journaled", where no action was taken, but the failures were recorded.

After running the SPC verification experiments, which were designed to test the SPC methodology, a series of production wafers were run without any problems. The only noticeable feature is a large spike in the later portion of the chart in Figs. 4 and 5 (Run#75 & #76). This was the result of a different material (Nitride deposited by a different process). Similarly the NU (Figs. 6 and 5) increased significantly. However, both ER and NU models were retuned after a small lag (Run #77) and the wafers for the remainder of the lot were all on target.

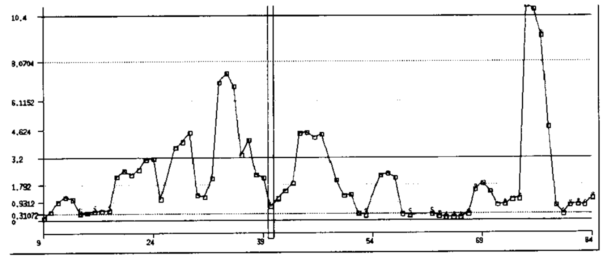


Fig. 8. Moving standard deviation chart for (actual - model) of etch rate.

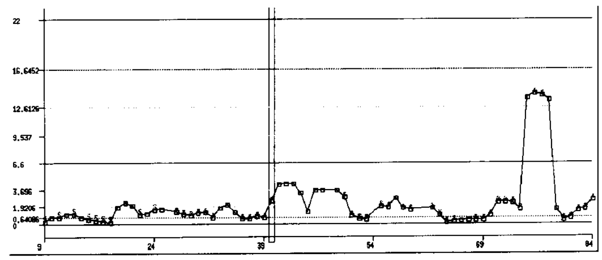


Fig. 9. Moving standard deviation chart for (actual - model) of nonuniformity.

V. Conclusion

This paper has demonstrated the general concepts, and the specific details, of operating a plasma etch process under the process control mode of operation. This mode allows the Operator to define the process not in terms of the control settings, but in terms of the requirements of all the measured observables. With sensors available for the in-situ measurement of these observables, an SQC procedure has been developed to monitor the (actual-predicted) values of these observables. When it is determined from SQC methods that the deviation from the anticipated result is statistically significant, the value of the constant term of the process/equipment model is updated in the original polynomial process model. An optimizer then automatically re-tunes the process, and provides a new recipe that comprehends the latest equipment state and considers all the observables, so as to keep all of them aimed at their respective target values.

To date, over 1000 wafers have been run in this mode of operation. As demonstrated in part in Figs. 4–9, each re-tuning of the process has vectored the system in the right direction, with the process observables tending closer to the predicted values. One of the interesting points to be made is that this type of model-based recovery is possible, even though the original model is no longer absolutely valid. Specifically, between the time of the original modeling work, and the implementation of this control algorithm, the baseline process (roughly at the center-point of the original experimental domain) had a significant decrease in ER from 60 to 50 A/sec. So while the original model is no longer valid, it still vectors the system in the proper direction when unexpected events drive the process astray.

ACKNOWLEDGMENT

The authors would also like to acknowledge the help of Bill Pu and Sonny Maung for implementing and debugging the process and machine control software. None of the experiments would be complete without the help of Daty Rogers who prepared the wafers, ran a part of the experiments, performed all the measurements and maintained AVP continually. Peter McAnally provided the standard PBL etch process.

VII. APPENDIX: CALCULATION OF SQC LIMITS

$$A_{\alpha/2} = \frac{1}{c_4}\frac{U_{1-\alpha/2}}{\sqrt{n}}, \tag{5}$$

$$B_{\alpha/2} = \frac{1}{c_4}\sqrt{\frac{\chi^2_{\alpha/2}}{\nu}}, \tag{6}$$

and

$$B_{(1-\alpha/2)} = \frac{1}{c_4}\sqrt{\frac{\chi^2_{(1-\alpha/2)}}{\nu}}; \tag{7}$$

where,

1) $[n]$ is the sample size,
2) $[\nu]$ is the degrees of freedom, and in this case $= n - 1$,
3) $[U_{\alpha/2}]$ is the value of the unit random variable with a normal probability distribution with a tail probability of $(1 - \alpha/2)$,
4) $[\chi^2_{\alpha/2}]$ is the value of the unit random variable with a chi-squared probability distribution with a tail probability of $(1 - \alpha/2)$,
5) $[\chi^2_{(1-\alpha/2)}]$ is the value of the unit random variable with a chi-squared probability distribution with a tail probability of $(\alpha/2)$, and
6) $[c_4]$ is the correction factor to account for the estimate of the population standard deviation from the sample standard deviation (it is a function of n).

For the $\overline{X}$ chart the SQC limits can be expressed as:

$$\begin{aligned} UAL &= \overline{X} + A'''_{.001} * s \\ UWL &= \overline{X} + A'''_{.025} * s \\ LWL &= \overline{X} - A'''_{.025} * s \\ LAL &= \overline{X} - A_{.001} * s \end{aligned}$$

For the s chart set parameters as

$$\begin{aligned} UAL &= B'_{.001} * s \\ UWL &= B'_{.025} * s \\ LWL &= B'_{.975} * s \\ LAL &= B'_{.999} * s \end{aligned}$$

For the case of 4 samples: the values of A, B coefficients are

$$\begin{aligned} A'''_{.001} &= 1.676 \\ A'''_{.025} &= 1.063 \\ A'''_{.975} &= -1.063 \\ A'_{.999} &= -1.676 \\ B'_{.001} &= 2.522 \\ B'_{.025} &= 1.911 \\ B'_{.975} &= 0.291 \\ B_{.999} &= 0.098 \end{aligned}$$

For individuals chart the action and warning limits are found by equations similar to that of the $\overline{X}$ chart, where the As are multiplied by $\sqrt{n}$ (n = sample size; 4 in this case). Along with these, the user has to specify a set of specification limits.

REFERENCES

[1] G. Taguchi, *Introduction to Quality Engineering: Designing Quality into Product and Processes.* UNIPUB. Kraus International Publications, White Plains, New York, 1986.

[2] G. Taguchi and M. S. Phadke, Quality Engineering Through Design Optimization. *IEEE Communication Society (Global Telecommunications Conference)*, Nov. 1984.

[3] G. E. P. Box, Studies in Quality Improvement: Signal to Noise Ratios, Performance Criteria and Statistical Analysis. Technical Report 11, Center for Quality and Productivity Improvement, University of Wisconsin-Madison, 1986.

[4] B. J. Bombay and C. J. Spanos, Application of Adaptive Equipment Models to a Photolithographic Process. In *SPIE Proceedings: Process Module Metrology, Control and Clustering*, volume 1594, pages 277 – 284, California, Sept. 1991.

[5] E. Sachs. The Run-by-Run Controller. In *SRC Workshop on Real Time Equipment Controllers*, Vancouver, Feb. 1991.

[6] Costas J. Spanos, Hai-Fang Guo, Alan Miller, and Joanne Levine-Parril. Real-Time Statistical Process Control Using Tool Data. *IEEE Transactions on Semiconductor Manufacturing*, pages 308 – 318, Nov. 1992.

[7] S. A. Henck. In Situ Real-Time Ellipsometry for Film Thickness Measurement and Control. *Journal of Vacuum Science Technology*, A 10(4):pp. 934, 1992.

[8] S. H. Naqvi, S. Gaspar, K Hickman, K. Bishop, and J. R. McNeil. Linewidth Measurement of Grating on Photomasks; A Simple Technique. *Applied Optics*, 1992:pp. 1377–1384, Apr. 1992.

[9] K. D. Allen, H. H. Swain, M. T. Mocella, and M. W. Jenkins. The Plasma Etching of Polysilicon with CF_3Cl/Argon Discharges. *Journal of The Electrochemical Society*, 133(11):pp. 2315, 1986.

[10] K. K. Lin and C. J. Spanos. Statistical Equipment Modeling For VLSI Manufacturing: An Application for LPCVD. *IEEE Transactions on Semiconductor Manufacturing*, 3(4):216–229, Nov. 1991.

[11] C. J. Davis, R. Matthews, and R. Hildenbrand. Vacuum Processing System. US Patent 4687542. Issu4ed on 10/24/85.

[12] P. K. Mozumder and G. G. Barna. Measurement of Spatial Nonuniformity of a Plasma Etch Process by the Slope of the Optical Emission Endpoint. To be published in IEEE Transactions on Semiconductor Manufacturing.

[13] J. W. Tuckey. *Exploratory Data Analysis.* Addison-Wesley, 1977.

[14] Bob Wheeler. *$ECHIP^{TM}$ Course Text.* ECHIP Inc., 724 Yorklyn Rd., Hockessin, DE 19707, 1989.

[15] George E. P. Box and Norman R. Draper. *Empirical Model Building and Response Surfaces.* Wiley series In Probability and Mathematical Statistics. John Wiley & Sons, New York, 1987.

[16] Philip E. Gill, Walter Murray, Michael A. Saunders, and Margaret H. Wright. *Users's Guide for NPSOL (Version 4.0): A Fortran Package for Nonlinear Programming.* Stanford University, Stanford, California 94 305, 1986.

[17] R. L. Iman and M. J. Shortencarier. *A FORTRAN 77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use With Computer Models.* Sandia National Laboratories, Albuquerque, New Mexico 87 185, 1984.

[18] Charles W. Champ and William H. Woodall. Exact Results for Shewart Control Charts With Supplementary Runs Rules. *TECHNOMETRICS*, 29(4):393–399, 1987.

[19] George E. P. Box, William G. Hunter, and J. Stuart Hunter. *Statistics For Experimenters.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1978.

[20] B. J. Mandel. The Regression Control Chart. *Journal of Quality Technology*, 1(1):pp. 1–9, Jan. 1969.

[21] D. C. Montgomery. *Introduction to Statistical Quality Control.* New York: Wiley, 2nd edition, 1990.

[22] J.I. Weindling, S. B. Littauer, and J. T. de Olivera. Mean Action Time of the $\overline{X}$ Control Charts with Warning Limits," *J. Quality Technol.*, vol. 2, no. 2, pp. pp. 79–85, Apr. 1970.

[23] S. Makridakis, S. C. Wheelwright, and V. E. McGee. *Forecasting: Methods and Applications.* New York: Wiley, 1983.

[24] D. C. Montgomery and L. A. Johnson. *Forecasting and Time Series Analysis.* McGraw Hill, New York, 1976.

[25] N. R. Draper and H. Smith. *Applied Regression Analysis.* Wiley series In Probability and Mathematical Statistics. John Wiley & Sons, New York, 1981.

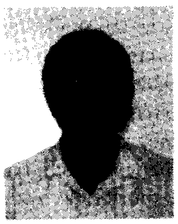
[26] John S. Oakland. *Statistical Process Control: A Practical Guide.* John Wiley & Sons Inc., New York, 1986.

[27] P. K. Mozumder and A. J. Strojwas. Statistical Control of VLSI Fabrication Processes. In S. W. Director, editor, *Proceedings of the IEEE: Special Issue on the Future of Computer-Aided Design*, volume 78, pages 436 – 455, Feb. 1990.

[28] R. B. Crosier. Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics*, 30(3):291 – 303, Aug. 1988.

[29] F. B. Alt. Multivariate Quality Control. In S. Kotz and N. L. Johnson, editors, *In Encyclopedia of Statistical Sciences*, volume 6, pages 110 – 122, New York, 1985. Wiley.

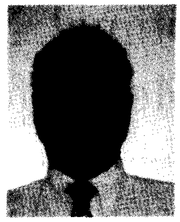
[30] W. H. Woodall and M. M. Ncube. Multivariate CUSUM Quality-Control Procedures. *Technometrics*, 27(3):285 – 292, Aug. 1985.



Purnendu K. Mozumder received his B. Tech Degree in electrical engineering from the Indian Institute of Technology in 1984, and his MS and PhD degrees in electrical and computer engineering in 1986 and 1989, respectively, from Carnegie Mellon University.

Dr. Mozumder is currently a member of the technical staff at the Semiconductor Process and Design Center of Texas Instruments Inc., Dallas, TX. He is presently working on design for manufacturing, TCAD, and statistical process control for VLSIC fabrication. He has been actively involved in the Microelectronics Manufacturing Science and Technology (MMST) project at TI. His interests include stochastic and adaptive control, statistical process and device simulation, TCAD Framework development, and application of statistical and numerical techniques to IC design and manufacturing methodologies.

Dr. Mozumder is presently serving as a guest editor of the IEEE Transactions on Semiconductor Manufacturing: Special Issue on MMST



Gabe Barna received hes B.Sc. (Honors Chemistry) and Ph. D. degrees from Mcgill University in Montreal, Canada in 1968 and 1973, respectively. After joining TI in 1972, he spent 11 years in the Central Research Laboratories working on the development of electrochemical sensors and electrochromic displays, LCDs and subsequently on the TI Solar Energy Project developing the fuel cell technology. In 1983, Gabe joined PAC, an organization that builds and supports plasma equipment within TI. There he worked on the hardware, software and process aspects of plasma processing. In the las 5 years has been involved in the Microelectronics Manufacturing Science and Technology (MMST) program, heavily forcused on the sensor-based control of plasma processes. He is a Senior Member of Technical Staff and has five patents and over 25 publications.