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
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 etching rate is uniform across the wafer. The controller can use the nitride etch rate as a feedback control.

The model-based process control operation of this reactor is illustrated in Fig. 1. In this reactor, the etching rate and uniformity are controlled by the model-based control strategy. The controller uses the nonuniformity of the etch as an important target for the process control strategy. The nonuniformity of the etch is measured using an in-situ sensor. The controller can then adjust the etching conditions to minimize the nonuniformity of the etch.

**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 the 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 output parameters required for the process control strategy described in this paper, ER (etch rate) is readily obtained together with the nonuniformity 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 (CHF3, CF4, O2, and Ar), the helium chuck pressure (He chuck), and the reactor 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.
of slope at the end-point were determined and then correlated to the \((\text{max.} - \text{min.})\) and \(\text{standard deviation} (\sigma/\mu)\) 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 \((\sigma/\mu)\).

Although linewidth (LW) control is significant for this etch, the \textit{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.

\textbf{Equipment Modeling}

With \textit{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\(^2\).

\(\text{Table I}\) 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\textsuperscript{TM} (a 31 wafer \textit{D-optimal} \cite{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.

\textbf{TABLE I}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
Parameter & Units & Low & Center & High \\
\hline
Pressure & mT & 100 & 150 & 200 \\
RF & W & 300 & 400 & 500 \\
\textit{CHF}_3 & sccm & 30 & 40 & 50 \\
\textit{CF}_4 & sccm & 40 & 60 & 80 \\
\textit{O}_2 & sccm & 10 & 15 & 20 \\
\textit{Ar} & sccm & - & 100 & - \\
\textit{He} - chuck & T & - & 2.5 & - \\
Substrate & °C & - & 0 & - \\
\hline
\end{tabular}
\caption{Input Parameter Ranges for PBL Model.}
\end{table}

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\textsuperscript{TM} \cite{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). These wafers were run past the nitride etch endpoint in order to determine the endpoint time under the individual process conditions.

\textbf{Set #1:}

\begin{itemize}
\item Substrate: Silicon
\item 1000Å
\item 2000Å
\item 2400Å
\item No Resist
\end{itemize}

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

\textbf{Set #2:}

\begin{itemize}
\item Substrate: Silicon
\item 1000Å
\item 2000Å
\item 2400Å
\item No Resist
\end{itemize}

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.

\textbf{Set #3:}

\begin{itemize}
\item Substrate: Silicon
\item 1000Å
\item 2000Å
\item 2400Å
\item CRB Positive Resist
\end{itemize}

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

\textsuperscript{2}Two line losses and two space gains were modeled. For the experiment, however, only one of the line loss models was used.
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 ECCHIP™, 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 (1.2µm 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 ($\bar{X}$) and

<table>
<thead>
<tr>
<th>Parameter</th>
<th>ER</th>
<th>NU</th>
<th>LW-1x1</th>
<th>LW-3x3</th>
<th>LW-1x5</th>
<th>LW-3x5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.951</td>
<td>0.927</td>
<td>0.947</td>
<td>0.901</td>
<td>0.894</td>
<td>0.905</td>
</tr>
<tr>
<td>Residual Std. Dev.</td>
<td>4.095</td>
<td>8.682</td>
<td>0.0106</td>
<td>0.0092</td>
<td>0.0181</td>
<td>0.0162</td>
</tr>
</tbody>
</table>
causes an increased lag between the time when an abnormal probability of false alarms. A larger value of 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 $\bar{X}$ and $s$.

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.

$$\hat{y} = f(x)$$  \hspace{1cm} (3)

$$y = \hat{y} + \epsilon$$  \hspace{1cm} (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) $f(x)$ is the equation of the model (the RSM for ER),
4) $x$ is the vector of input parameters, and
5) $\epsilon$ is the error term, that signifies the noise in the prediction error.

The term $\epsilon$ is the key parameter for performing model based SQC. The theory of regression states that $\epsilon$ 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

---

3The Individual ($X$) charts were not used for SQC to minimize the probability of false alarms.
normally with $N(0, \sigma)$ assuming insignificant lack of fit [15], where $\sigma$ 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 $\epsilon$. This forms the basis of the SQC charts, with mean value of 0.0 and standard deviation of $\sigma$ [20], [21]. An unbiased estimate of $\sigma^2$ 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 is more susceptible to false alarms than the $\bar{X}$ 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 SQC algorithms is provided in Section IV. It is also understood that using the $X$ 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 $X$ 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 $\bar{X}$ chart trends follow 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 $X$ chart. Both these effects would cause false SQC failures and make the controller oscillate to instability.

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 factor of 0.50 (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 $X$. 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 $X$ 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 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 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 O2 Calibration Table also misadjusted so that O2 flow was 60% of the menu value.
6) W-17 O2 Calibration Table again misadjusted so that O2 flow was 50% of the menu value.
7) W-18 O2 Calibration Table again misadjusted so that O2 flow was 200% of the menu value.
8) W-26 RF Limit table and MFC calibration values reset to correct values.

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 (X), and Moving Standard Deviation (s). Each chart shows the cumulative run number on the abscissa, and the individual values or statistical moments of the deviations between the model prediction and actual values of the outputs, 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 NU were 50Å/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, the larger the time lag before the WECO rules triggers. This
* 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.

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.
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.

The next significant discontinuity (Run#20) was seen for wafer W-10 where misadjusting 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 2σ 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

The S 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.

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 10secm but only about 5secm 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 O2 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
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 (Run#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 60Å/s to 45Å/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 O₂) 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 8.8%. 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.

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 Å/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 Maun 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

\[ A_{\alpha/2} = \frac{1}{c_4} U_{1-\alpha/2} \sqrt{n}, \]

\[ B_{\alpha/2} = \frac{1}{c_4} \sqrt{\chi^2_{\alpha/2}/\nu}, \]

and

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

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 \((\alpha/2),\)
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 \(\bar{X}\) chart the SQC limits can be expressed as:

\[ UAL = \bar{X} + A_{9901} \times s, \]
\[ UWL = \bar{X} + A_{9255} \times s, \]
\[ LWL = \bar{X} - A_{9255} \times s, \]
\[ LAL = \bar{X} - A_{9901} \times s. \]

For the \(s\) chart set parameters as

\[ UAL = B_{9901} \times s, \]
\[ UWL = B_{9255} \times s, \]
\[ LWL = B_{9255} \times s, \]
\[ LAL = B_{9901} \times s. \]

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

\[ A_{9001} = 1.676, \]
\[ A_{9255} = 1.063, \]
\[ A_{9900} = -1.676, \]
\[ B_{9001} = 2.522, \]
\[ B_{9255} = 1.911, \]
\[ B_{9900} = 0.291, \]
\[ B_{9999} = 0.098. \]

For individuals chart the action and warning limits are found by equations similar to that of the \(\bar{X}\) chart, where the \(s\) 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


**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 manufacturability, 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 his 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 last 5 years has been involved in the Microelectronics Manufacturing Science and Technology (MMST) program, heavily focused on the sensor-based control of plasma processes. He is a Senior Member of Technical Staff and has five patents and over 25 publications.