Real time statistical process control for improved etch tool performances

Gene Smith
Cypress Semiconductor, 3901 N. First Street, San Jose, California 95135

Ivar Highberg
35949 Nicolett Ct., Fremont, California 94536

(Received 3 May 1995; accepted 20 September 1995)

Through the use of a commercially available computer integrated manufacturing software package, it has been shown that the efficiency of a batch reactive ion etch reactor has increased dramatically. Standard statistical process control methodology has been applied to process input parameters in eliminating the need to run periodic "process qualification tests," identify lot-to-lot variations, and to improve the scheduling of preventative maintenance functions. These techniques have shown an increase in the first pass yield, by reducing scrap, and have proven invaluable in both process development and optimization. This has in turn increased the capacity and extended the capability of the tool. These techniques are straightforward and can be applied to other tools with potentially similar results. © *1996 American Vacuum Society.*

I. INTRODUCTION

The etcher discussed within is an Applied Materials' 8300 series hexode (batch) reactor used to etch multilayer metal film stacks. An additional 486 personal computer is required for the computer integrated manufacturing (CIM) software package (PCHost83™). This also has an ethernet interface to connect to our network environment. An optical barcode wand is attached to the PCHost83 computer to allow product material or lot identification through access to the inventory tracking database residing on the network.

The Highberg Associates software is an interface program that resides in the PCHost83 computer connected to the etcher. It has several functions including automatic recipe downloading (through barcoding), data acquisition of any number of the *in situ* process variables, and keyboard macros that allow many functions to be strung together allowing unattended operation for many hours. These macros are used to process material, perform various calibration functions, and when required, help troubleshoot the system.

The computer network is a local area network (LAN) that connects the inventory control and process data database on a mainframe computer to the etcher through the PCHost83 computer (Fig. 1). The PCHost83 computer is a buffer between the etcher and the main frame network where all communications to and from the network go through it.

The operator first logs the production material into the specific production activity (etch) using the optical barcode reader. This information is then sent through the network to the work in progress (WIP) inventory tracking database. Any given lot of material is then recognized by the WIP tracking database and cross referenced with computerized specifications. The WIP function notifies the PCHost83 computer of the proper process recipe, which in turn orders the etcher to run that process recipe. The operator then loads wafers into the etcher and starts it. The PCHost83 keyboard macro function initiates the wafer transport (one or two cassettes) and the process recipe functions. At the same time it gathers process data for transfer to various databases on the network. This procedure has removed the possibility of misprocessing due to improper recipe selection by the operator.

II. DATA TYPES AND STORAGE

A multitude of *in situ* process data can be captured from the etcher. This includes (but is not limited to): the radio-frequency (rf) bias voltage generated during the process, the process pressure, the rf power generated, the emission intensity levels throughout the process, the process temperature, the throttle valve position, etch time by recipe step, and mass flow controller (MFC) flow by recipe step. Machine events ("door closing," "load lock pump down complete," etc.) and system alarms ("cannot reach desired dc bias," "throttle

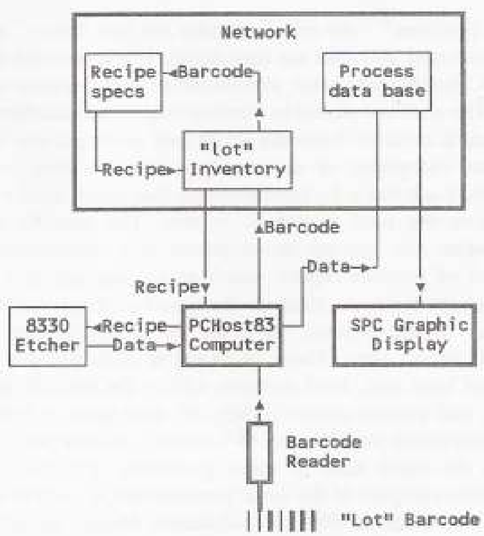

FIG. 1. PCHost83 computer network block diagram highlighting the barcode identification of material, automatic recipe downloading, data acquisition, and display.

 0734-211X/96/14(1)/533/5/$10.00

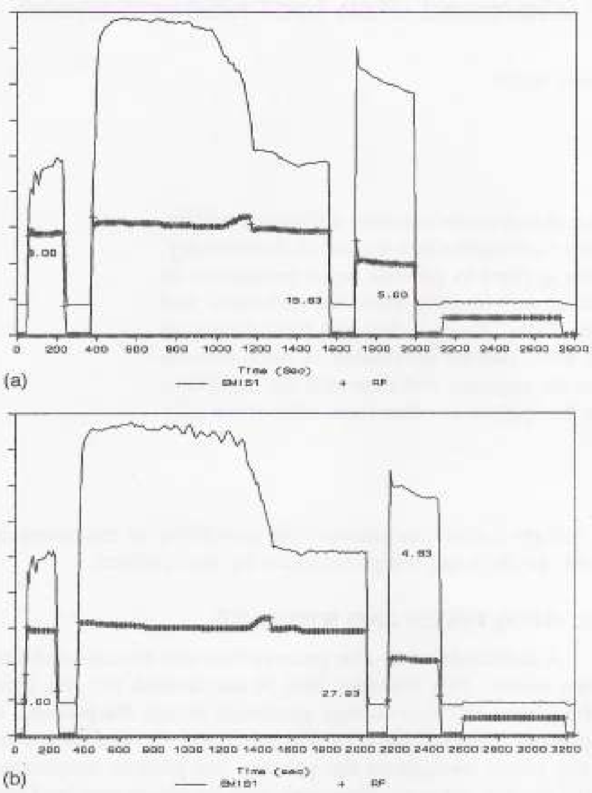

FIG. 2. (a) Typical endpoint trace for a "sandwich" metal film stack. Substep etch times (minutes) are highlighted. (b) Endpoint trace for the same metal film stack as (a) but with a different mask density. Note the total etch time is significantly longer for the second substep.

valve position," "the robot encoder has lost home," etc.) are also captured. All data are transferred to the network through the PCHost83 computer immediately after process completion. The data are stored in various structured database tables that are accessible from the 8330 tool administrator's desk.

Four categories of data analysis can be done with this system. Each has a focused purpose that when used together, comprise the total analytical system. The specific analysis categories are: process development and optimization, SPC control of process inputs, material accounting, and system error/alarm analysis. Each is discussed in detail below.

Any of the process variables mentioned above can be plotted versus time. These plots are used to monitor the effects of load size, load position within the reactor, substrate types, and pattern density. They are also used in optimizing overetch times which reduces substrate consumption and reduces the cycle time of those processes. Figures 2(a) and 2(b) are examples of the same process recipe used to etch the same film stack on different substrates. Notice the difference in the emission signal and the time that the aluminum took to endpoint. This analysis showed the need to segregate these two devices. Prior to this work, both devices were etched together. One can easily see how one device could be underetched while the other overetched.

Developing new processes for different metal film stacks is accomplished first by creating a base line of critical parameters for each type of device to be processed. Substrates, mask density, and pitch, along with the load size all contribute to process variability. By looking at and accounting for each variability individually, one can control the reactor to much tighter limits.

The data can be plotted utilizing SPC methods to analyze the day-to-day control of the reactor. The critical process variables can be viewed in the following four ways: raw data, first differences, moving average, and cumulative sums.

Raw data [Fig. 3(a)] is the data as retrieved and plotted against time for each boat (not lot) of material processed. Control limits are then calculated based upon the last several months of data. First differences [Fig. 3(b)] can also be called the first derivative and represents the graphing of data points obtained by the following formula:

$$\sum_{1}^{n} (B_n - B_{n-1}).$$

B is the data for one boat load of wafers processed while n is the number of boat loads plotted and used to calculate the control limits. It is good to take the effect of time out of this data and look at the rate of change.

One can see that for the time to endpoint data, the first difference is valuable in identifying boat-to-boat variances. This was the metric to measure the effect of changes to the operation of the etcher with respect to load sizes, chamber conditions, and any preventative maintenance techniques.

Since lots of material are generally larger than the 8330's 18 wafer maximum load, it requires 2 passes through the reactor to process 1 lot. Viewing the average of two lots [Fig. 3(c)] is helpful. This is shown by plotting the data obtained using the following formula:

$$\sum_{1}^{n} \frac{B_n + B_{(n-1)} + B_{(n-2)} + B_{(n-3)}}{4}.$$

This approach smoothes out the data and makes it easier to interpret.

Cumulative sums [Fig. 3(d)] are shown by plotting the data obtained using the following formula:

$$\sum_{1}^{n} \text{all data} - B_n ,$$

where all data is the average of all the data plotted and B is the specific data collected for one boat of wafers.

Figure 3(d) shows the cum-sums results for rf power delivered for each boat load of material processed. This approach enables the identification of a significant difference as indicated by changes in the slope of the data. Note that there was a significant change in the data every time there was a preventive maintenance (PM) performed on the system with the exception of one time. This was when the rf generator

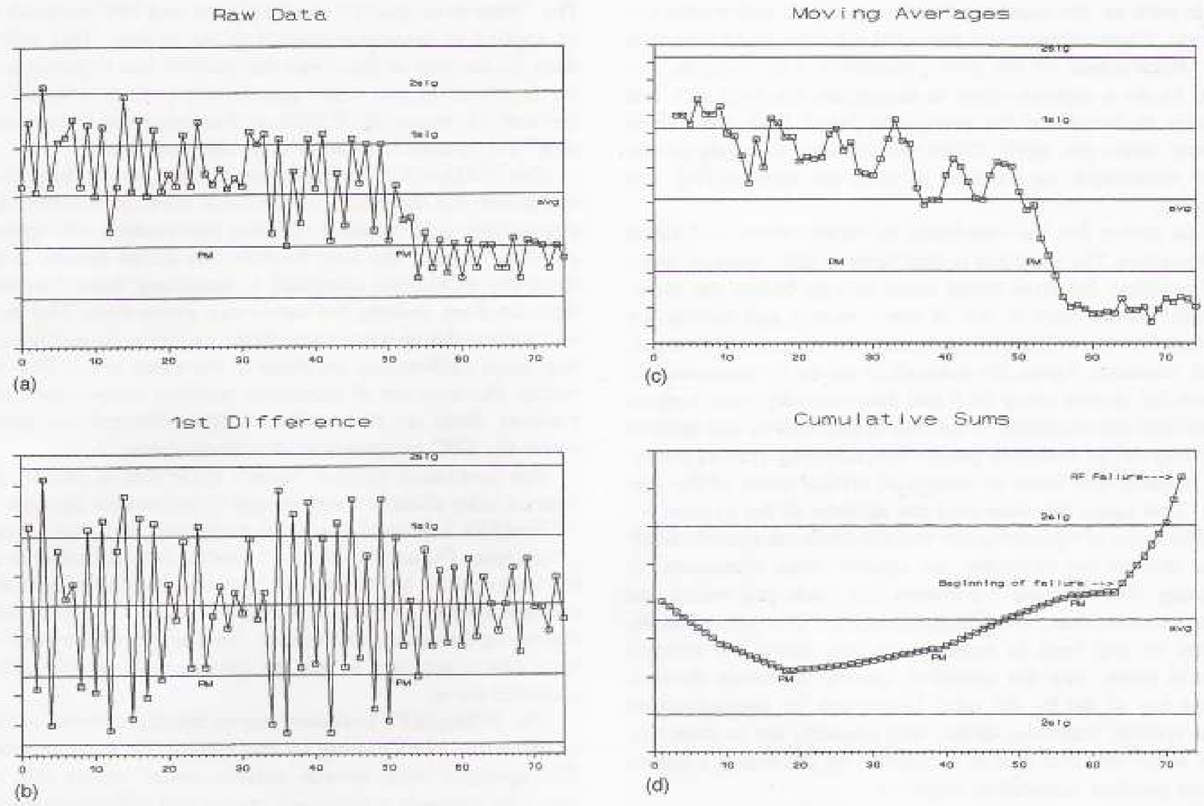

FIG. 3. (a) Example of the raw "time-to-endpoint" data. Each data point is the time it took for one load of material to reach endpoint. The time-to-endpoint changes with the preventative maintenance functions (highlighted). (b) Example of the first difference of the same raw time-to-endpoint data in Fig. 3(a). Notice the change at each preventative maintenance function. (c) The same raw data from (a) was plotted using moving averages of 4. This method helps smooth out the noisy data to see trends not normally seen. Notice the data has a downward trend for the first 50 data points but is pretty flat after the second preventative maintenance function. (d) The rf watts delivered to the chamber is plotted using cumulative sums. Each data point is one load of material. Cum-sums is helpful by looking at the change in slope. Note that at each preventive maintenance function (highlighted) the slope changed with the exception of the last slope change. This data shows when the rf generator actually began to fail.

actually started to fail. The system failed several boat loads later. But, this system detected the beginning of the end of this component.

All data obtained is automatically analyzed every 24 h to determine whether any critical process parameter has drifted out of control (OOC). Should this happen, a report to the tool administrator is electronically mailed with the specific data relating to the OOC point. This report contains all pertinent data relating to which parameter is out of control, the number of data points used to generate the control limits, how other critical parameters relate, and the date and time the violation occurred. For every report generated, an assignable cause can be determined by viewing the plots shown in Figs. 2 and 3. All of this analysis can be performed by the engineer at his desk. Monitoring the same plots periodically has helped to identify potential problems about to occur. This is also efficient since only periodic review of the data is required to automatically determine if everything is under control. The data can also be used to tighten the control limits to further identify and reduce variability.

III. MATERIAL ACCOUNTING AND ALARM ANALYSIS

All data obtained from product material processed is analyzed using basic pareto techniques. Items reviewed are the number of passes, the number of wafers processed, the type of recipes used and how often, the number of rf hours (total and per recipe substep) on a per week basis.

This data is utilized to schedule chamber wet cleans based upon use, which avoids the age old struggle of manufacturing's desire to schedule preventative maintenance (PM) based upon days versus the engineer's desire to schedule PM based upon use. Since there are several different metal film stacks being etched with both Cl_x and F_x chemistries, the reactor conditions will change when one recipe gets used more often than another. Changes in reactor conditions can affect the control of the system. The "by recipe substep" data serves as a history of the recipes used and helps identify changes in reactor conditions. Additional calculations are made such as: the number of wafers per pass and wafers per rf hour. These calculations provided a metric to demonstrate the effectiveness of the new procedures. For example, we used to do a separate pass to determine the etch rate and another to determine the particulate level. Both were done several times per week. When the separate etch rate passes were eliminated, an increase in units per hour (UPH) was realized.

The etcher has the capability to collect event and alarm information. The problem is that there is little storage space at the etcher. An error could come and go before the maintenance owner sees it. All of these events and alarms are captured by the PCHost83 computer and transferred to a network database. Again, the network is set up to automatically search for alarms every 24 h and electronically mail a report to the tool administrator. This data is then stored and utilized [in analysis] to maintain proper functionality, predict potential failures, and focus resources on critical areas of the system. This again has increased the up time of the system.

The cycle of tightening the control limits on specific hardware devices by changing the system setup (constants) is ongoing. The PCHost83 software has made this easier and more accurate than conventional methods. The etcher has the ability to flag (and in some cases shut down the reactor) several errors that are aimed at specific hardware devices. These are all set by the user. Limits can be determined on each specific hardware device and properly set to detect issues at the slightest hint of a problem by predicting a failure before product material is impaired.

Historically, the reactor was periodically subjected to a nonproduction load of material that was used to determine the etch rate of each process. Additional time was spent to verify the MFC calibration (rate of rise methodology) as well as a separate pass to determine particulate levels. All of these add up to time not available to production as well as increased intervention by the operator. Both affect the UPH of the system. The need to run qualification passes was eliminated because the same information could be determined by calculating and monitoring etch rates in normal production (see Figs. 2 and 3).

Etch rates (E/Rs) were usually plotted on a SPC chart to determine the control of the reactor representing an "output response" that consists of several hardware "input variables." If the E/R is out of control, it generally takes several hours to determine the particular subsystem at fault, fix it and requalify. Why not control the specific "input" hardware devices in question?

Using the data analysis techniques discussed above, various hardware components' performances can be plotted and SPC rules applied. The methodology of monitoring the process inputs as opposed to the process outputs is much more accurate and generally points to the specific failing device before the process fails. An added benefit is that since the data is automatically collected, this analysis is "free" and thereby very cheap to implement.

The customary aspect of measuring a flat substrate, etching it and rereading it to determine E/Rs is a relic of the past. The "time-to-endpoint" can be plotted and SPC methods can be applied to determine control of the system. This will reduce the number of functions the operator has to perform and the overhead of test wafer generation; both of which lower the cost of ownership (COO) and increase the "time available" the system is to process product material.

The PCHost83's keyboard macro function can be utilized to perform the chamber leak rates and MFC flow calibrations sequentially with minimal operator intervention (the operator still has to push the start button). An added benefit is that since the PCHost83 computer is executing these functions, they are done exactly the same way every time. This again will reduce the operator to operator variation by making sure that these calibrations are done in the same order. This will reduce the time out of service to perform maintenance calibrations. Both are then automatically collected and plotted under the SPC environment described above.

The customary plasma "burn" cycle that is usually performed after chamber wet cleans is automated through the PCHost83's keyboard macro to minimize time and operator interaction. Dummy wafers are loaded into the reactor and the leak checks, MFC flow calibrations, and the plasma burn are executed in succession without separate load-lock pump downs or operator intervention. This again will reduce the time that is required to get the system back on line after a chamber clean.

The PCHost83's keyboard macro function can be utilized to allow trouble shooting of the reactor for particle issues. The operator loads several particle count wafers into one boat and executes a keyboard macro that will cause successive wafers to be loaded and processed one at a time. Each wafer has a different process that flows a different gas. This enables the operator to obtain particle data on all gases with a single operation. The PCHost83 computer is performing these functions ensuring they will be done exactly the same every time. This removes any doubt about the testing technique and allows one to focus on specific problems faster. This also increases the up time of the system.

Since specific system event reports have a time and date stamp associated with them, they can be looked at to evaluate various subsystem's performance. For example, the cryopump regeneration takes a long time to perform. That activity was historically scheduled on a calendar basis, not on use. By simply monitoring the time the pump is turned on until it reaches total pump down provides an indicator of when it is necessary to perform the regeneration. This is also applied to the other roughing pumps.

IV. RESULTS

When one adds up all facets of this automated network, the time available to production has increased nearly 15%. This was done simultaneously while implementing several requirements of tighter linewidths and lower substrate consumption budgets. This in turn has allowed the reactor to process more wafers with better capability, thereby pushing out additional capital equipment requirements. The reduction of human interaction through automated routines has also contributed to the time available as well as extending the scope of what one operator or maintenance technician can do while maintaining tighter control. The increase in up time has alone paid for the PCHost83 software and the computer in less than six months not to mention the increased control which has lead to better performance and an increased capability of the reactor.

V. SUMMARY

Through the use of a commercially available CIM software package "networked" to the inventory tracking computer database, and basic SPC methodology of "process inputs," the up time has been increased an estimated 15% for an Applied Materials' 8330 reactor. This has increased control of the reactor and actually increased the capability of it by allowing it to function on tighter and tighter geometries while reducing substrate and resist consumption budgets. An added benefit has been reduced scraps through automated recipe selection and decreased operator intervention; all of which have reduced the cost of ownership and extended the life of the tool. The methods employed can be implemented in a short time and for minimal cost (to some degree or another) based upon network hardware and capability.