Real-time virtual metrology and control of etch rate in an industrial plasma chamber

Shane A. Lynn¹, Niall MacGearailt², and John V. Ringwood¹

Abstract—Plasma etch is a semiconductor manufacturing process during which material is removed from the surface of semiconducting wafers, typically made of silicon, using gases in plasma form. A host of chemical and electrical complexities make the etch process notoriously difficult to model and troublesome to control. This work demonstrates the use of a real-time model predictive control scheme to control plasma etch rate in the presence of disturbances to the ground path of the chamber, which are representative of maintenance events. Virtual metrology (VM) models, using plasma impedance measurements, are used to estimate the plasma etch rate in real time for control, with a view to eliminating the requirement for invasive measurements. The VM and control schemes exhibit fast set-point tracking and disturbance rejection capabilities. Etch rate can be controlled to within 1% of the desired value. Such control represents a significant improvement over open-loop operation of etch tools, where variances in etch rate of up to 5% can be observed during production processes due to disturbances in tool state and material properties.

I. INTRODUCTION

High-volume, high-yield, and high-throughput manufacturing is of primary importance in modern semiconductor manufacturing. Product wafers in a semiconductor manufacturing cycle typically undergo over 350 different process steps in their path from raw silicon wafer to finished product.

Plasma etch is a key process in the manufacturing cycle. During plasma etch, etchant gases in plasma form are directed towards the wafer surface using electric and magnetic fields. The gases react with the exposed areas of the wafer surface, and the etch product subsequently evaporates to remove material. Plasma etch is conducted within specialised etch chambers. Process input variables to the chambers are typically well controlled variables such as chamber pressures, component temperatures, and gas flow rates, that are specified by set points. In general, the required etch process input variables are developed through extensive experimentation and the etch recipes are typically applied to product wafers in an open-loop manner [1].

Etch processes exhibit process drift and unpredictable shifts in behaviour due to chamber conditioning, incoming wafer variability, and the unpredictable effects of chamber maintenance activities. In-situ measurements of etch rate and etch depth are expensive, time consuming, and often invasive. Measurements are not available to machine operators without a considerable metrology delay and it is impractical for fabrication plants to measure every wafer processed. Significant quantities of wafer scrap can result if a tool operates out of specification undetected costing manufacturers significant revenue unnecessarily.

Plasma etch processes are predominantly managed using statistical process control (SPC) [2], where variables measured in-situ during each process, or variables concerning the result of each process, are monitored for deviations that indicate erroneous operation. However, metrology delays can still result in wafers being processed erroneously, and the process recipe does not take the typical time-varying nature of the process into account.

Advanced process control (APC) and virtual metrology (VM) are enabling technologies that can resolve the control issues in semiconductor manufacturing [3]. APC includes fault detection, fault classification, fault prognosis, and process control. APC implementation in the semiconductor industry has broadly been restricted to lot-to-lot control [4] because of infrequent measurements and large metrology delays. VM is the timely estimation of process metrology variables that may be expensive or difficult to measure using readily available process information as shown in Fig. 1.

![Virtual metrology principle](image)

This paper focusses on real-time control of plasma etch rate. Plasma etch rate is typically regulated using wafer-to-wafer control schemes, potentially using VM for feedback [5]. Real-time control of plasma etch rate has been reported by a number of researchers, but typically requires the use of bulky, expensive, or invasive measurement techniques. For example, Sarfaty et al. [6] use laser reflectance interferometry (LRI), Stokes and May [7] use data from LRI, laser interferometry (LI), residual gas analysis (RGA), and optical emission spectroscopy (OES), and Rosen et al. [8] use in-situ spectroscopic ellipsometry measurements of wafer thickness to implement real-time control.
This work investigates the feasibility of real-time VM, using plasma impedance monitor (PIM) data, for real-time control of plasma etch rate. The VM estimates are obtained non-invasively, allowing cost-effective, real-time control without process perturbation. Control of etch rate is achieved using model-based predictive control (MPC) with applied RF power as the manipulated variable. Control is implemented using predictive functional control (PFC), an MPC technique.

Preventative maintenance (PM) events have a large influence on etch performance and VM model estimation accuracy. PM events involve the routine replacement of components, such as electrodes and ceramic covers, that have been exposed to etchant chemicals for possibly over 1000 wafer etch and cleaning cycles. Although the replacement components are macroscopically identical to those that are removed from the chamber, microscopic differences in the electrical connections made between components when they are replaced change the electrical characteristics of the chamber. At the high frequencies in use during plasma processing (≈ 13.56 MHz), changes in impedance, stray capacitances, and stray inductances cause considerable changes to the electrical behaviour of the chamber and hence the etching plasma properties [9]. The electrical path between the powered chamber electrode and ground (the ground path) influences plasma variables such as the ion flux to the etching wafer and the DC bias of the wafer in the chamber [10]. Hence, changes in the impedance of the ground path brought about by PM events can cause the etch performance of the chamber to vary dramatically across maintenance cycle events. For the experiments described in this paper, an extra modified match box that allows manual control of impedance is installed on the ground path from the chamber. Hence, variations in the ground impedance can be realised as required, partially simulating the effect of PM events. The ground impedance variations act as unmeasured disturbances to the plasma, changing plasma variables such as electron density, and affecting the etch performance.

In this paper, Section II describes the experimental apparatus used. Section III provides the theoretical background for PFC. Section IV describes the development of VM models for etch rate. Finally, Sections V and VI give the control results and paper conclusions, respectively. The results in this paper, along with further results on the real-time control of plasma electron density [11], have been submitted for publication in journal format in [12].

II. APPARATUS

An overview of the experimental setup is shown in Figure 2.

Plasma is generated in a capacitively-coupled, top-powered, parallel-plate plasma etch chamber. Between 0 and 625 W of RF power at 13.56 MHz is delivered to the topmost chamber electrode from an RF generator. Chamber pressure is controlled to a specified set point by means of a gate valve between the etch chamber and a vacuum turbo pump. The bottom electrode in the etch chamber is grounded through a modified match unit, such that the position of the matching inductor can be varied manually, effecting a total ground impedance of between 0 – 70 Ω. Variations in this path act as disturbance signal to the plasma in the chamber. The plasma undergoes a mode change at approximately 25Ω, above which etching at the wafer surface stops. Hence, for the experiments described here, the ground impedance is limited to a ceiling of 25Ω. This range of ground impedance encapsulates those changes that typically occur as a result of PM events.

A plasma impedance monitor (PIM) is an electronic sensor that is installed between the matching network and the plasma electrodes. The PIM sensor provides information on the current, voltage and phase of the waveforms on the power supply circuitry. Information on the fundamental frequency of 13.56 MHz and up to 52 harmonics of this frequency is recorded. Power at the harmonic frequencies of the supplied power is generated in the supply circuitry as a result of the non-linear impedance presented by the oscillating plasma sheaths at the electrodes in the plasma chamber. Calculations of impedance, reactance, resistance, and power can be made from the PIM signals.

Two PIM sensors are used. One PIM is installed on the powered electrode of the chamber, and provides information on the applied RF power. The second sensor records information about the path to ground from the chamber. Analog output channels on each PIM sensor provide real-time measurements which can be used for control. Unfortunately, the analog signals undergo a fixed delay of 0.5 seconds as a result of the ADC circuitry used by the PIM processing units.

The VM input variables are measured using a National Instruments USB-6009 ADC connected to the PIM sensors, and can be sampled at a higher rate than the electron density.
probe. For the experiments in this work, the analog sampling rate (and hence the VM measurement rate) is set to 10 Hz to allow for noise averaging during each sample. All of the control calculations are performed using the MATLAB® programming environment.

For etching experiments, 200 mm diameter silicon wafers coated with a thin layer (approximately 6000 Å) of polysilicon are used. To measure the true average etch rate of the process, the thickness of the polysilicon layer is measured precisely using an interferometer at 19 different locations on the wafer surface both before and after wafer processing. The mean depth change over all of the measured locations is recorded as the etch depth achieved for the wafer. The average etch rate is calculated by dividing the etch depth by the known etch time. Real-time measurements of etch depth are not available.

III. PREDICTIVE FUNCTIONAL CONTROL

A. Motivation

Predictive functional control (PFC) is a MPC technique where the internal models used are independent internal models that depend solely on the process input. The manipulated variable is constructed on a set of basis functions, typically a polynomial basis [13]. PFC is chosen for the control of plasma etch because it is easily implemented using a first-order approximation to the system, it uses a single intuitively interpreted parameter during tuning, it is designed primarily for single-input single-output systems, and it can control the etch system taking the VM delay into account.

B. Internal model

The “internal model” is a model of the plant used by a predictive controller that is capable of predicting future process outputs. Consider a zero-order hold equivalent model for a first-order process, having a gain \( K_m \) and a time constant \( \tau_m \) as

\[
y_m(k) = a_m y_m(k-1) + b_m K_m u(k-1),
\]

where \( a_m = e^{-\tau_m/T_s} \), \( b_m = 1 - a_m \), and \( T_s \) is the system sample period. The actual process output is \( y_p \). Equation (1) describes an independent model that calculates the output \( y_m \) using only the known measured process inputs and past model outputs. Because the process may be subjected to unknown disturbances and the plant model will not be perfect, \( y_p \neq y_m \). However, \( y_p \) and \( y_m \) will evolve in parallel, and the model is used to calculate increments of the process output rather than the absolute response of the process subjected to a particular input [13].

The model prediction of the process response to a step change in input, from the instant \( k = 0 \) to a future time \( k + H \), where \( H \) is an integer number of samples, consists of the sum of the free solution \( y(k)a_m^H \) and the forced solution \( K_m u(k)(1 - a_m^H) \).

C. Reference trajectory

The desired future behaviour of the controlled variable is the “reference trajectory”. The reference trajectory is initialised on the current process output \( y_p(k) \), and defines the path taken by the controlled variable to the current set point \( S(k) \). The “coincidence horizon” is the set of points in the future where the process and the model outputs should be equal. For the sake of simplicity, only one coincidence point \( H \) is considered. Typically, an exponential reference trajectory is defined such that the error signal at a time \( k + H \)

\[
S - y_p(k + H) = e(k + H) = e(k)\lambda^H,
\]

where \( S \) is a constant set point, \( \lambda = e^{-\tau_r} \), with \( \tau_r \) the desired closed-loop time constant of the controlled system.

The desired process output increment at the coincidence point, \( \Delta y_p(k + H) \), is given by

\[
\Delta y_p(k + H) = e(k) - e(k + H).
\]

Hence

\[
\Delta y_p(k + H) = -e(k)\lambda^H + e(k) = (S - y_p(k))(1 - \lambda^H).
\]

At each sample \( k \), the values for \( \Delta y_p \) are computed, and the first value is applied to the plant and model. At sample \( k + 1 \), the procedure is repeated, resulting in a new reference trajectory, in essence creating a receding horizon.

D. Calculation of controlled variable

The new manipulated variable \( u(k) \) is expressed as a weighted sum of \( N \) basis functions that are chosen according to the nature of the process and set point variations:

\[
u(k + i) = \sum_{j=0}^{N-1} \mu_j F_j(i), \quad 0 \leq i \leq H,
\]

where \( F_0, F_1, \cdots, F_{N-1} \) are the basis functions, and \( \mu_0, \mu_1, \cdots, \mu_{N-1} \) are the weights associated with the functions. PFC generally uses a set of polynomial basis functions i.e., \( F_j(i) = i^j \). In the elementary case, including the case that applies here, the basis functions reduce to \( N = 1 \), \( F_0(i) = i^0 = 1 \). Thus

\[
\Delta y_m(k + H), \quad \text{where} \quad \Delta y_m \quad \text{is the model increment, is given by}
\]

\[
\Delta y_m(k + H) = y_m(k + H) - y_m(k),
\]

\[
\Delta y_m(k + H) = y_m(k)a_m^H + K_m u(k)(1 - a_m^H) - y_m(k).
\]

By setting \( \Delta y_p(k + H) = \Delta y_m(k + H) \) we can solve for the manipulated variable \( u(k) \)

\[
u(k) = \frac{(S - y_p(k))(1 - \lambda^H) - y_m(k)a_m^H + y_m(k)}{K_m(1 - a_m^H)}.
\]

The main tuning parameter becomes the desired closed loop response time (CLRT), which is specified by \( \tau_r \). The controller is tuned by adjusting the value of \( \tau_r \), the desired closed-loop time constant. Although MPC-based controllers can control a system with zero steady state error in the presence of model mismatch [14], the model will incorrectly
estimate the required process input increments at each sample with the result that the CLRT will not exactly match the desired time constant \( \tau_r \).

### E. Systems with a pure time delay

For the work described here, the 0.5 s delay in the VM estimates acts as a pure time delay equivalent to five sample periods in the system. The delay is not included in the PFC internal process model so that, ideally for a delay of \( d \) samples, \( y_p(k) = y_m(k-d) \), and \( y_{predict}(k+d) = y_p(k+d) = y_m(k) \). Hence, the change in the process output between times \( k \) and \( k + d \) is equal to the change of the model output between times \( k - d \) and \( k \), yielding

\[
y_p(k + d) - y_p(k) = y_m(k) - y_m(k - d),
\]

which can be solved for \( y_p(k + d) \). Hence, the reference trajectory is not initialised on the current of value \( y_m(k) \), but on the predicted value of \( y_p(k + d) \) in order to anticipate its response. The control equation given in (8) is still valid by replacing \( y_p(k) \) with the expression for \( y_p(k + d) \).

### IV. VIRTUAL METROLOGY

For VM modelling, data are collected from the system excited over an experimental range, and, following this, an empirical VM model relating the plasma etch rate to the recorded PIM data is constructed. During the experiments, the pressure in the etch chamber is kept constant at 300 mTorr, a typical production process pressure.

Three modelling techniques known to produce accurate estimates for plasma etch modelling [1], [15], Multiple linear regression (MLR) [16], back-propagation artificial neural networks (ANNs) [17], and Gaussian process regression (GPR) [18] models, are examined as candidate empirical modelling techniques for VM. The ANNs used have a single hidden layer that is varied in size from one to fifteen neurons and randomly initialised five times during model training. These configurations are typically seen as the best-performing for plasma manufacturing processes [1]. The GPR models use a squared exponential covariance function.

To gather data for VM models, 19 wafers were etched in a He/SF\(_6\) plasma at 300 mTorr using different power levels and ground impedance values. The ground impedance was varied from 0 – 25 \( \Omega \) at powers of 200, 300, and 400 W, etching one wafer for each combination of input values. Etch depth for each wafer was measured after processing, and the corresponding etch rates were calculated. 6 extra wafers were etched at intermediate settings of power and ground impedance to test the VM models. The thickness of the polysilicon layer on the wafer is measured before and after etching and the average etch rate is calculated.

The input data used by the VM models are taken from the upper PIM sensor, comprising the fundamental values of the powered electrode current, voltage, phase, and the calculations of the plasma impedance, reactivity, resistance, and power. Stepwise regression is used to determine the PIM variables that best explain the variance in the etch rate values across the training data set. Current, phase, and power are highlighted as significant variables. Table I shows the VM accuracy achieved by the modelling techniques investigated. Models are compared in terms of mean absolute percentage error (MAPE). An adequate fit for estimation of etch rate was achieved using linear regression models as seen in Figure 3.

<table>
<thead>
<tr>
<th></th>
<th>Training MAPE (%)</th>
<th>Test MAPE (%)</th>
<th>Max Test Error (( A/\text{min} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLR</td>
<td>0.56</td>
<td>0.50</td>
<td>55.8</td>
</tr>
<tr>
<td>ANN</td>
<td>0.34</td>
<td>0.68</td>
<td>68.4</td>
</tr>
<tr>
<td>GPR</td>
<td>0.00</td>
<td>1.21</td>
<td>122.5</td>
</tr>
</tbody>
</table>

**TABLE I**

VM ESTIMATION RESULTS. \( R^2 \) VALUES FOR ALL MODELS ARE GREATER THAN 0.99.

Fig. 3. Etch rate estimates from linear regression model.

### V. ETCH RATE CONTROL

Real-time, in-situ measurements of plasma etch rate are unavailable during control experiments. As a result, only the average etch rate over each complete wafer run can be used to validate the control scheme accuracy. Prior to the implementation of the PFC control scheme, a proportional-integral (PI) controller which facilitates intuitive manual tuning is first investigated as a potential control solution for the plasma etch rate.

Problems with PI control arise from as a result of the 0.5 second delay caused by the PIM measurements. The delay causes integral windup when the PI controller first starts. Additionally, control is further complicated by the initial transients of the system when the plasma is first ignited. In the initial seconds of the discharge, the molecules in the etchant gases disassociate, resulting in an instantaneous increase in volume and correspondingly, chamber pressure. There is a finite time delay (~ 3 seconds) before the chamber control system counteracts this pressure change to recover the 300 mTorr pressure set point. During this transient period, the controller attempts to compensate for the etch rate values reported by the VM model by increasing the power to the
ER Target | Ground imp. | ER achieved | Error
(Å/min) | Ω | (Å/min) | (%)
--- | --- | --- | ---
4500 | 23 | 4511 | 0.2
4500 | 24.8 | 4508 | 0.2
4500 | 13.5 | 4500 | 0.0
4500 | 14.5 | 4488 | 0.3
5500 | 14.5 | 5513 | 0.2
5500 | 16.8 | 5556 | 1.0
5500 | 22 | 5438 | 1.1
5500 | 24.5 | 5362 | 2.5

**TABLE II**

*Etch rate (ER) control using PI control with VM system. The MAPE over all experiments was 0.2%.*

To attenuate potentially dangerous fluctuations in the applied power, a 2.5 second delay time is implemented at the start of each control period, during which the applied power is fixed and the controller is disabled. To accurately control etch rate during this startup phase of plasma etch, the VM model for etch rate would be required to include this dynamic scenario. Although such a model may improve the control results, the model would be infeasible to develop due to the complexities of the different etchant gases and etch chambers that are encountered.

With the startup safeguard in place, the PI controller is capable of regulating etch rate based on the real-time VM estimates relatively accurately. To test the etch rate control accuracy, wafers are etched with set points in the etch rate that are validated after processing. Unmeasured disturbances are introduced to the ground impedance prior to the etch of each wafer, partially simulating the effect of disruptive PM events. Table II gives the etch rate results for PI control for a number of different etch rate targets and ground impedance values. Larger errors are reported for values with high ground impedance values and powers because the plasma starts to change mode in this operational state and the etch rate VM model becomes less accurate.

For PFC, the system, at a specified pressure, can be approximated as a pure gain \( K_m \), with negligible dynamics and a delay term such that

\[
G_m(s) = K_m e^{-\tau_d s}
\]  

(10)

where \( \tau_d = dT_s \) is the VM delay in seconds. No dynamics are used because the relationship between power and plasma etch rate is virtually instantaneous. The lack of dynamics in the system model simplifies the PFC equations since \( a_m = e^{-\tau_m} = 0 \). Equation (8) reduces to

\[
u(k) = \frac{(S - y_p(k))(1 - \lambda_H) + y_m(k)}{K_m}
\]

(11)

Five wafers were etched with the PFC control scheme to a target etch rate of 3000 Å/min. The PFC control scheme was tuned such that \( K_m = 16 \), as determined from the relationship between the applied power and the VM estimate of etch rate, and \( \tau_r = 1 \) sec. The time constant of 1 s is chosen as it is relatively small compared to a typical production step time (60-80 seconds approx.), and transitions at this rate will not significantly affect final etch results. With \( T_s = 0.1 \) s, ten samples are taken per time constant, allowing reliable control. We observed that with smaller values (\( \tau_r \sim 0.1 \) s), the control system becomes sensitive to noise on the VM estimates, resulting in noisy controlled variables. The ground impedance was varied randomly for each wafer. Table III shows the results of this test. As shown in Figure 4, the PFC control scheme results in more desirable startup transients than those of the PI controller.

As shown in Figure 5, the VM and control scheme adjusts the applied power in response to natural within-wafer variance of the chamber conditions to maintain a consistent etch rate. The ground impedance remained constant during this test. It is hypothesised that the changing etch rate is caused by gradual temperature changes of the chamber during the etch process. Such within-wafer control is advantageous for process reliability when compared to more typical open-loop or run-to-run control schemes where the applied power remains constant for the duration of each wafer etch.

Finally, the real-time VM scheme offers a method to estimate the etch depth during the etch process. Figure 6 shows the estimated etch rate for a wafer where the etch rate set point has been changed during the wafer run (using PI control). The etch depth reported by the interferometer after the wafer was processed was 4727.1 Å. Integrating the VM etch rate estimate over the experiment duration estimates the etch depth with 0.81% error at 4689.0 Å. Hence, it is feasible.
Fig. 5. Real-time adjustment of applied power to maintain a consistent average etch rate in response to within-wafer process variance.

to provide a real-time etch depth estimate by integrating the VM etch rate estimate during the etch process.

VI. CONCLUSIONS

This research has demonstrated the feasibility of real-time VM and control schemes for plasma etch rate in an industrial environment. The use of PIM sensors for VM allows control without perturbation of the plasma, and relatively minor modifications to chamber hardware. The continuous adjustment of process inputs in real-time, during wafer etch, represents a marked improvement upon existing manufacturing techniques, where process inputs are only adapted on a per-wafer or per-lot basis.

However, there are limitations to the technique that require further investigation. Firstly, the VM models may become invalid due to process drift over long durations [19]. When real metrology from drifting processes is available, the VM model will require refreshing to maintain model currency (as proposed in [4]). A similar refreshing technique may be required for the independent model of the PFC system, as investigated in [12]. Secondly, separate VM models may be required for different tools in fabrication plants due to mismatch between individual tools and sensors. Finally, further investigation to characterise the effects of PM events is required to ensure the VM models can remain accurate.

VII. ACKNOWLEDGEMENTS

The authors would like to thank the staff of Intel Ireland Ltd. for their help with the plasma etch chamber and help with the computer systems of the fabrication environment. This work was supported in part by the Irish Research Council for Science, Engineering, and Technology.

REFERENCES