Real-time Virtual Metrology and Control of Plasma Electron Density in an Industrial Plasma Etch Chamber ⋆

Shane A. Lynn * **Niall Macgearailt** ** **John. V. Ringwood** *

* *Dynamics and Control Group, Electronic Engineering Department, National University of Ireland, Maynooth (Tel: +353 1 7086907; e-mail: shane.a.lynn@eeng.nuim.ie).*
** *School of Electronic Engineering, Dublin City University (DCU), Dublin, Ireland (e-mail: niall.macgearailt@eeng.dcu.ie)*

Abstract: Plasma etching is a semiconductor manufacturing process during which material is removed from the surface of silicon wafers 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 maintain a consistent plasma electron density in the presence of disturbances to the ground path of the chamber. The electron density is estimated in real time using a virtual metrology model based on plasma impedance measurements. Recursive least squares is used to update the controller model parameters in real time to achieve satisfactory control of electron density over a wide operating space.

1. INTRODUCTION

Plasma etch is a fundamental process used in semiconductor manufacture in which etchant gases are excited to plasma form and used to remove precise amounts of material from exposed surfaces of silicon wafers. Although considerable research has been undertaken in the area of estimation and control of etch processes (Ringwood et al. (2010)), plasma etch processes are still typically run in semiconductor fabrication plants in an "open-loop" fashion, where etch recipes are specified in terms of process inputs such as powers, gas flows, and pressures, in quantities that are known to produce the desired etch performance. However, specification of etch recipes in this manner is prone to error due to three principal causes:

- Chamber performance changes over time as the chamber becomes conditioned by repeated etch operations.
- Chamber maintenance operations, carried out at specific intervals, lead to unpredictable shifts in etch performance as parts are replaced and/or cleaned.
- The behaviour of process hardware such as mass flow controllers (MFCs) and matching networks drifts over time.

The resulting unpredictable changes in etch performance prevent reliable process reproducibility, and statistical process control (SPC) is typically used to monitor the etch process (May and Spanos (2006)). A more ambitious approach is to specify etch recipes in terms of plasma variables such as electron density, electron temperatures, radical densities, and ion fluxes to the wafer surface.

Real-time control of plasma variables is required to achieve such recipe specification, enabling precise control of the etch process, and reducing sensitivity to process drift and disturbances. For example, Hankinson et al. (1997) regulate etch rate through real-time control of fluorine concentration and wafer bias voltage by manipulating chamber power and pressure. Lin et al. (2009) control electron density measured using a transmission line microstrip microwave interferometer to eliminate the "first-wafer effect" from experiments, that is the tendency of the first wafer from each lot processed to exhibit peculiarities in etch performance. Klimecky et al. (2003) demonstrate real-time electron density control to reduce etch rate variances without affecting the etch profiles of product wafers. Direct measurements of the plasma variables were used in the control schemes.

Real-time measurement of controlled variables often requires bulky, expensive, or invasive equipment. *Virtual metrology* (VM) is the use of mathematical models to estimate variables that may be difficult or expensive to measure using readily available process information. In this work, a VM system is employed to estimate plasma electron density in real time using readily available plasma impedance data. Hence, no invasive sensors that perturb the etch process are required, and the VM model is used to provide electron density estimates for feed-back control purposes.

A model predictive control algorithm, predictive functional control (PFC), is employed to control the electron density, and the internal model of the controller is updated in real time using recursive least squares (RLS) regression. As shown in Fig. 1, the process inputs and the estimated process outputs from the VM model are used to recursively update the PFC internal model parameters.

Variations in the ground path impedance (the electrical path between the chamber electrode and ground) are in-

⋆ This work was supported in part by the Irish Research Council for Science Engineering and Technology (IRCSET).

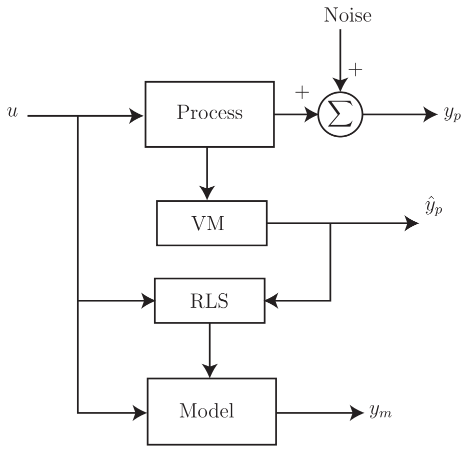

Fig. 1. Real-time model identification using RLS.

troduced as disturbances to the plasma chamber. Step changes in the ground impedance are postulated to reflect the type of change that can occur when chamber components such as electrodes and ceramics are replaced or cleaned during maintenance events.

The aim of this work is to regulate the plasma electron density without the use of additional invasive sensors in the plasma chamber. The ability to compensate for disturbances caused by maintenance operations has the potential to reduce etch rate variability, wafer scrap, and tool downtime in manufacturing environments.

2. EXPERIMENTAL EQUIPMENT

Fig. 2 shows the apparatus used to implement the real time virtual metrology and control system.

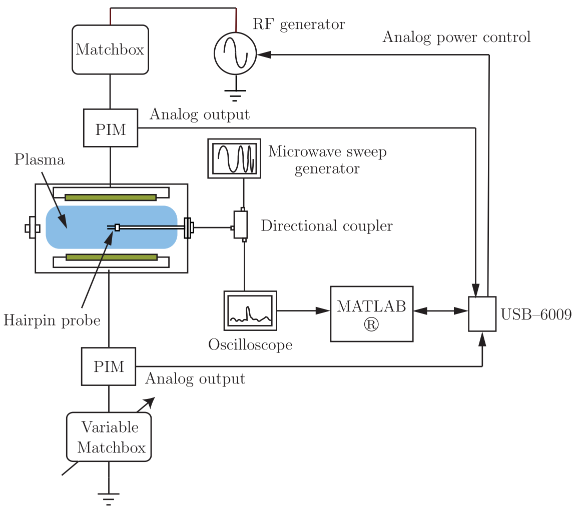

Fig. 2. Virtual metrology and control hardware.

2.1 Plasma etch chamber

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. The amount of delivered power is specified via a $0-10$ V reference signal generated using a control computer using DAC hardware. Chamber pressure is controlled to specified set points by means of a gate valve between the etch chamber and the 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 $0-25\ \Omega$. Variations in this path act as disturbance signal to the plasma in the chamber.

2.2 PIM sensors

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. 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 for control.

2.3 Hairpin resonator probe

The electron density in the plasma etch chamber is determined using a microwave hairpin resonator, or simply *hairpin probe*. Introduced by Stenzel Stenzel (1976) in the mid 1970's, a hairpin probe is an open-ended quarter wavelength transmission line whose resonant frequency is related the dielectric constant of the medium that surrounds it. Figure 3 shows a schematic of a hairpin probe.

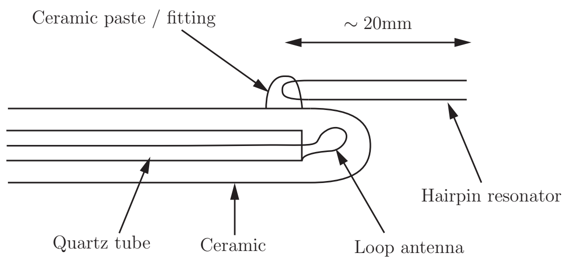

Fig. 3. Microwave resonator hairpin probe.

The plasma electron density is related to the frequency difference between the hairpin resonances with and without the plasma,

$$n_e = \frac{f_r^2 - f_0^2}{0.81}, \tag{1}$$

where $n_e \times 10^{10}$ cm^{-3} is the electron density, and f_r and f_0 are the resonant frequencies (in gigahertz) of the hairpin with and without the plasma respectively (Piejak et al. (2004)).

3. VIRTUAL METROLOGY OF ELECTRON DENSITY

The microwave probe is an invasive measurement of plasma electron density that presents a number of disadvantages if used for control:

	Low	High
Chamber power (W)	200	600
Ground Impedance (Ω)	0	22
Pressure (mTorr)	200	300

Table 1. Design of experiment inputs for VM model with varying pressure.

(1) Production wafers cannot be etched while the probe is inserted in the chamber as the plasma is perturbed around the probe body.
(2) The sampling frequency of the probe is limited to 2 Hz due to the time required to download and process the reflected current waveform from the oscilloscope.

As indicated in Section 1, virtual metrology (VM) is used to estimate the electron density for control purposes. The dynamics of the etch chamber system are virtually instantaneous, and as such, static VM models are employed.

The VM models are created empirically from data collected from the chamber. The chamber parameters are varied over the ranges specified in Table 1 while accompanying electron density measurements are recorded for model training purposes. Training and test data sets are collected for creating the VM models and testing their generalisation performance, respectively. The input variables used by the VM models are measurements taken from the upper PIM sensor, comprising the fundamental values of the powered electrode current, voltage, phase, and the fundamental values of the plasma impedance, reactance, resistance, and power. The VM input variables are measured using an 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.

Multiple linear regression (MLR) (Montgomery et al. (2001)), artificial neural networks (ANNs) (Haykin (1999)), and Gaussian process regression (GPR) (Rasmussen (1996)) 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. The GPR models use a squared exponential covariance function. The modelling results are summarised in Table 2 where an ANN model is the most accurate VM model over the unseen test data. MLR models perform worst due to the non-linear relationship that exists between the VM input variables and the electron density. The performances of the ANN and GPR models are quite similar, with the ANN models performing better on unseen test data. Offsets between the estimated and real values of electron density are observed for some system operating points. However, these offsets are rarely greater than 1×10^9 cm^{-3} ($\sim 2-3\%$ absolute error), which is deemed an acceptable level of error for the experimental control work.

The VM scheme introduces a delay in the control system. VM estimations are delayed by a constant value of 0.5 s compared to the actual electron density as a result of the operation of the internal circuitry in the PIM sensors. This time delay must be catered for in the control paradigm implemented.

	Training MSE	Test MSE	Max Test Error
MLR	2.612	2.512	5.943
ANN	1.004	0.870	3.632
GPR	0.675	1.210	3.345

Table 2. VM estimation results. R^2 values for all models are greater than 0.99.

4. PREDICTIVE FUNCTIONAL CONTROL

4.1 Motivation

Model predictive control (MPC) or model-based predictive control (MBPC) was first employed in the 1970's the defence and petroleum industries. Predictive functional control (PFC) is differentiated from the other forms of MPC in that the internal models used are independent internal models that depend solely on the process input. Furthermore, the manipulated variable is constructed on a set of basis function, typically a polynomial basis (Richalet and Donovan (2009)).

4.2 Internal model

The "internal model" is a model of the plant used by a predictive controller that is capable of predicting future process outputs. The internal model is not restricted to a particular form and can be formulated as a transfer function, state-space, step-response, black-box model etc. Consider a first-order process with a gain K_p and a time constant τ_p subject to an input u. The plant is represented by the following difference equation:

$$y_p(k) = a_p y_p(k-1) + b_p K_p u(k-1), \tag{2}$$

where $a_p = e^{\frac{-T_s}{\tau_p}}$, $b_p = 1 - a_p$, and T_s is the sampling time. The plant can be modelled by a first order system with a gain K_m and a time constant τ_m as

$$y_m(k) = a_m y_m(k-1) + b_m K_m u(k-1), \tag{3}$$

where $a_m = e^{\frac{-T_s}{\tau_m}}$, $b_m = 1 - a_m$. Equation 3 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 (Richalet and Donovan (2009)).

The prediction of the process response using only the process model from the instant $k = 0$ to a future time $k + H$, where H is an integer number of samples, consists of the free solution $y(k)a_m^H$ and the forced solution $K_m u(k)(1 - a_m^H)$. By superposition, the full solution $y(k+H)$ is the sum of the free and the forced responses.

4.3 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 $k + H$ is

$$S - y_p(k + H) = e(k + H) = e(k)\lambda^H, \qquad (4)$$

where S is a constant set point, $\lambda = e^{\frac{-T_s}{\tau_r}}$, and τ_r is the required closed-loop time constant of the controlled system. The controller is tuned by adjusting the value of τ_r which is easily interpretable.

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

$$\Delta p(k + H) + e(k + H) = e(k) \qquad (5)$$

Hence

$$\Delta p(k + H) = -e(k)\lambda^H + e(k) = (S - y_p(k))(1 - \lambda^H) \qquad (6)$$

At each sample time k, the values for Δp are computed, and the first value is applied to the plant and model. At the next sample time, $k + 1$, the procedure is repeated, resulting in a new reference trajectory, and in essence creating a sliding horizon.

4.4 Calculation of controlled variable

The future manipulated variable $u(k)$ is structured around a set of basis functions that are chosen according to the nature of the process and set point variations:

$$u(k + i) = \sum_{j=0}^{N-1} \mu_j F_j(i), \; 0 \le i \le H. \qquad (7)$$

Thus the manipulated variable is expressed as a weighted sum of N basis functions. PFC uses a set of basis functions that consist of a polynomial basis, i.e., $F_j(i) = i^j$. In the elementary case, and the case that applies here, the basis functions reduce to $N = 1$, $F_0(i) = i^0 = 1$.

As seen before $\Delta p(k+H) = (S - y_p(k))(1-\lambda^H)$. $\Delta m(k+H)$, Δm is the model increment, is given by

$$\Delta m(k + H) = y_m(k + H) - y_m(k), \qquad (8)$$

$$\Delta m(k + H) = y_m(k)a_m^H + K_m u(k)(1 - a_m^H) - y_m(k). \qquad (9)$$

The equation $\Delta y_p(k + H) = \Delta y_m(k + H)$ is fulfilled by

$$(S-y_p(k))(1-\lambda^H) = y_m(k)a_m^H + K_m u(k)(1-a_m^H) - y_m(k), \qquad (10)$$

which can be solved for the manipulated variable $u(k)$

$$u(k) = \frac{(S - y_p(k))(1 - \lambda^H) - y_m(k)a_m^H + y_m(k)}{K_m(1 - a_m^H)}, \qquad (11)$$

This is the fundamental PFC control equation in its most elementary form (Richalet and Donovan (2009)).

4.5 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 in the system. Predictive controllers can take this time delay into account. 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

$$y_p(k + d) - y_p(k) = y_m(k) - y_m(k - d) \qquad (12)$$

which rearranges to

$$y_{predict}(k + d) = y_p(k + d) = y_p(k) + y_m(k) - y_m(k - d). \qquad (13)$$

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 (11) is still valid by replacing $y_p(k)$ with the expression for $y_{predict}(k + d)$ in (13).

Many processes in production industries can be approximated by a first order system model, and in many PFC control applications, an exponential reference trajectory is used with a single coincidence horizon point $H = 1$ and a zero order basis function. Hence the main tuning parameter becomes the desired CLRT which is specified by τ_r.

4.6 Plasma chamber model

The relationship between the power delivered to the chamber electrode and the plasma electron density is approximately linear for constant values of ground impedance and pressure as indicated by Fig. 4. Considerable changes in system gain occur when the chamber pressure is changed and for each pressure set point, smaller changes in gain are observed as the ground impedance of the chamber is altered.

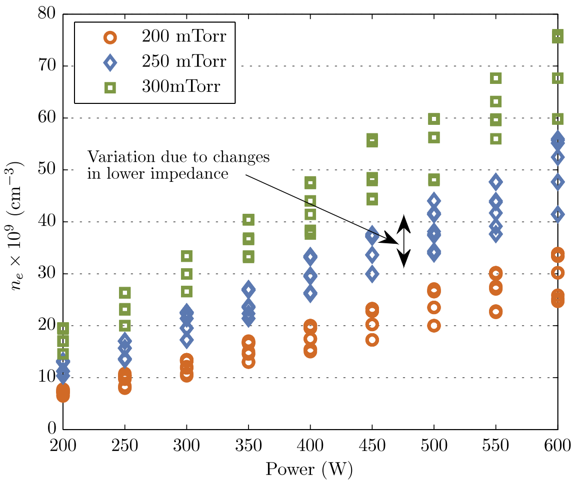

Fig. 4. Electron density response to power at different pressures.

The system, at a specified pressure, can be approximated as a pure gain K_m, with negligible dynamics and a delay term such that

$$G_m(s) = K_m e^{-\tau_d s} \qquad (14)$$

where $\tau_d = dT_s$ is the VM delay in seconds. No dynamics are used in this model because the relationship between power and electron density is instantaneous. The lack of dynamics in the system model simplifies the PFC control equations since $a_m = e^{\frac{-T_s}{\tau_m}} = 0$ and the system model equation without delay will consist of the forced solution alone. Equation (11) reduces to

$$u(k) = \frac{(S - y_p(k))(1 - \lambda^H) + y_m(k)}{K_m}. \quad (15)$$

Although MPC-based controllers can control a system with zero steady state error even in the presence of model mismatch (Rossiter and Richalet (2001)), the model will incorrectly estimate the required process input increments at each sample with the result that the closed-loop time constant will not match the desired time constant τ_r. To achieve accurate control over a range of pressures, and hence a range of system gains, recursive least squares is used to update the model parameters in real time.

5. REAL-TIME MODEL IDENTIFICATION

Due to the relatively simple form of the system model in use, the model update can be constructed as a linear regression problem. At each sample moment k, the electron density VM measurement $\hat{n}_e$ is given by

$$\hat{n}_e(k) = K_p p(k-d) + c_m + \xi \quad (16)$$

where K_p is the process gain, p is the power applied to the chamber, d is the system delay in sample periods, c_m is an offset term, and ξ is zero-mean Gaussian white noise. This linear equation has a non-zero $\hat{n}_e$-intercept, included in the model as an input disturbance. For samples k to $k+N$,

$$\begin{bmatrix} \hat{n}_e(k) \\ \hat{n}_e(k+1) \\ \vdots \\ \hat{n}_e(k+N) \end{bmatrix} = \begin{bmatrix} p(k-d) & 1 \\ p(k-d+1) & 1 \\ \vdots & \vdots \\ p(k-d+N) & 1 \end{bmatrix} \begin{bmatrix} K_m \\ c_m \end{bmatrix}. \quad (17)$$

Equation (17) is of the form $\mathbf{y} = X\boldsymbol{\beta}$ which can be solved using MLR techniques. Rather than storing a fixed window of inputs and outputs and recalculating new values for K_m and c at each sample, the well-known recursive least squares (RLS) technique can be used to update $\hat{\boldsymbol{\beta}} = [K_m c_m]^T$ at each sample.

The RLS algorithm can be implemented in five steps at each sample time, $k+1$ (see Wellstead and Zarrop (1991)):

(1) Form row vector $\overrightarrow{\mathbf{x}}(k+1)$, made up of the latest measurement from the process $[p(k-d+1)\ 1]$.
(2) Calculate the current error $e(k)$ using
$$e(k+1) = n_e(k+1) - \overrightarrow{\mathbf{x}}(k+1)\hat{\boldsymbol{\beta}}(k). \quad (18)$$
(3) Calculate the covariance matrix $\mathbf{P}(k+1)$
$$\mathbf{P}(k+1) = \mathbf{P}(k)\left[I - \frac{\overrightarrow{\mathbf{x}}(k+1)^T\overrightarrow{\mathbf{x}}(k+1)P(k)}{\lambda_{RLS} + \overrightarrow{\mathbf{x}}(k+1)\mathbf{P}(k)\overrightarrow{\mathbf{x}}(k+1)^T}\right] \quad (19)$$
where $0 \leq \lambda_{RLS} \leq 1$ is a forgetting factor.
(4) Update the model parameters,
$$\hat{\boldsymbol{\beta}}(k+1) = \hat{\boldsymbol{\beta}}(k) + \mathbf{P}(k+1)\overrightarrow{\mathbf{x}}(k+1)^T e(k+1). \quad (20)$$
(5) Return to step 1.

6. CONTROL RESULTS

Satisfactory control is achieved at a constant pressure without the inclusion of the RLS correction for the PFC internal model. Experimental results for set point tracking, and set point tracking in the presence of disturbances in ground impedance, are shown in Figs. 5 and 6 respectively. While the controller sample time is $T_s = 0.1$ s, the electron density is measured separately at sample rate $T_s = 0.5$ s. The ANN VM model accurately estimates the electron density in real time. Although the specified closed-loop time constant for both experiments is $\tau_r = 1$ s, the system actually responds with a time constant of approximately 1.6 s due to a mismatch between K_m and K_p.

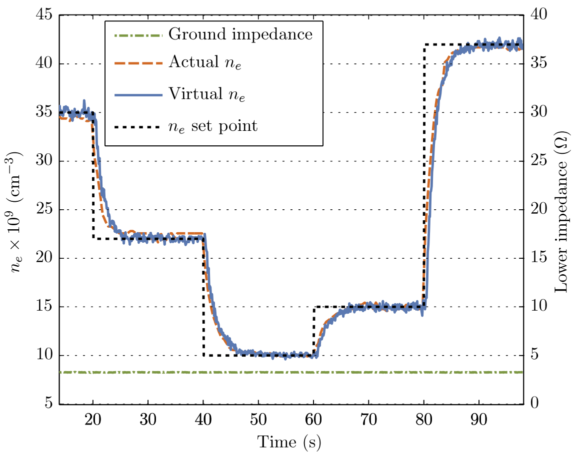

Fig. 5. PFC control of electron density with $T_s = 0.1$ s, $\tau_r = 1 = 1$ s, at constant pressure.

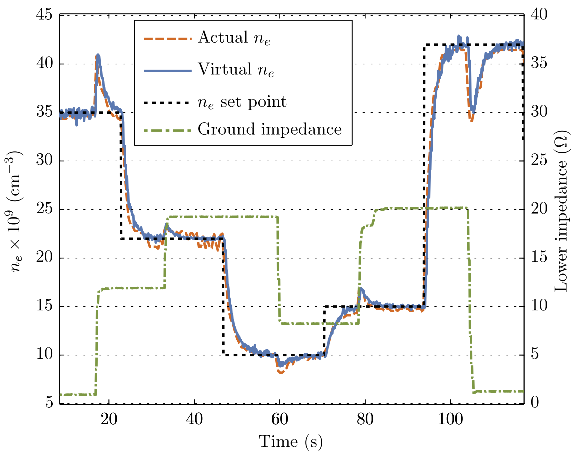

Fig. 6. PFC control of electron density with $T_s = 0.1$ s, $\tau_r = 1$ s, at a constant pressure in the presence of unmeasured disturbances in ground impedance.

To demonstrate the effectiveness of the RLS model update, Fig. 7 shows the control achieved by the PFC controller using a constant value of K_m over a range of pressures. K_m is set to a value that suits higher pressure regions, and the plasma chamber undergoes two step changes in pressure and ground impedance, as indicated by the annotations on the figure. The model output is higher than the plant estimated output for the duration of the experiment due to the model/plant gain mismatch. Slower transients than those specified by τ_r are observed. The effect of the model/plant gain mismatch lessens as the chamber pressure is increased and the plant gain rises. Fig. 8 demonstrates the effect of the RLS adaptive scheme.

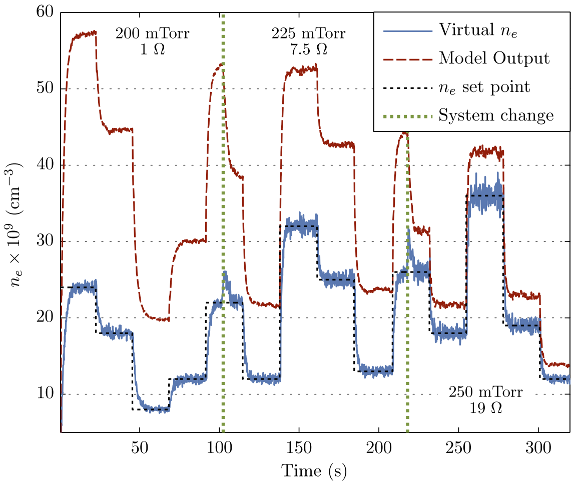

Fig. 7. Effect of pressure disturbances on PFC controller performance.

The model gain is adapted to suit the plant input and estimated output at each sample point. A relatively slow forgetting term of $\lambda_{RLS} = 0.995$ is employed so that the model updates do not react to noise in the system, but adapt slowly as samples become available. The transient responses, after model convergence, match the required time constant τ_r much more accurately than those shown in Fig. 7. However, directly after system changes before the model parameters have been updated to the correct values, the system can suffer from overshoot in cases where $K_m < K_p$ and slow responses when $K_m > K_p$.

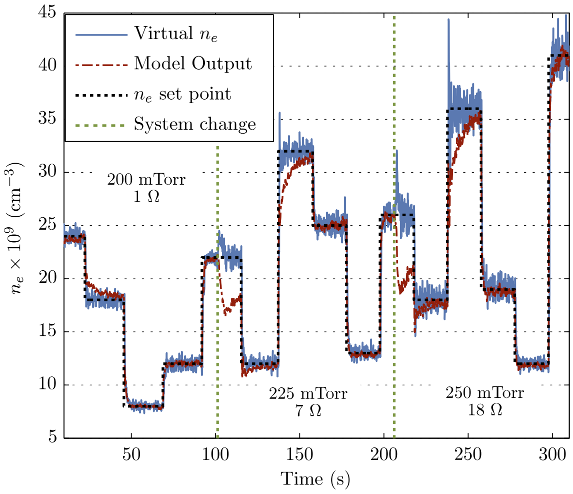

Fig. 8. PFC results using internal model realigned by RLS with pressure changes.

7. CONCLUSIONS AND FUTURE WORK

Control of plasma electron density is achieved using a first-order PFC implementation, achieving relatively fast set point tracking (settling times of approximately 1 second) without overshoot and excellent disturbance rejection properties. The control system is expanded to operate across a large operating range by adapting the PFC internal model parameters using RLS.

The electron density VM and control scheme is currently being expanded to operate during wafer etch operations using production gases and a range of chamber conditions. Experiments to construct VM models in the presence of layered silicon wafers are being carried out. When completed, the control scheme will reduce process variability caused by chamber maintenance operations and drift by adjusting recipe power settings to achieve a desired plasma electron density, thereby enabling more reliable and consistent etch performance.

ACKNOWLEDGEMENTS

The authors would like to thank David Kavanagh of Dublin City University for his assistance in this work.

REFERENCES

M. Hankinson, T. Vincent, K.B. Irani, and P.P. Khargonekar. Integrated real-time and run-to-run control of etch depth in reactive ion etching. *IEEE T. Semiconduct. M.*, 10(1):121–130, Feb. 1997.

S. Haykin. *Neural Networks: A Comprehensive Foundation.* Prentice Hall, 1999.

Pete I. Klimecky, J. W. Grizzle, and Fred L. Terry. Compensation for transient chamber wall condition using real-time plasma density feedback control in an inductively coupled plasma etcher. *J. Vac. Sci. Technol. A.*, 21(3):706–717, May. 2003.

C. Lin, K.C. Leou, H.M. Huang, and C.H. Hsieh. Feedback control of plasma electron density and ion energy in an inductively coupled plasma etcher. *J. Vac. Sci. Technol. A.*, 27(1):157–164, Jan. 2009.

G. S. May and C. J. Spanos. *Fundamentals of semiconductor manufacturing and process control.* Wiley-Interscience, 2006.

D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to Linear Regression Analysis.* John Wiley & Sons, Inc., 2001.

R. B. Piejak, V. A. Godyak, R. Garner, B. M. Alexandrovich, and N. Sternberg. The hairpin resonator: A plasma density measuring technique revisited. *J. Appl. Phys.*, 95(7):3785–3791, Apr. 2004.

C. E. Rasmussen. *Evaluation of Guassian processes and other methods for non-linear regression.* PhD thesis, Department of Computer Science, University of Toronto, Toronto, ON, Canada, 1996.

J. Richalet and D. O. Donovan. *Predictive Functional Control.* Springer, May. 2009.

J. V. Ringwood, S. Lynn, G. Bacelli, B. Ma, E. Ragnoli, and S. McLoone. Estimation and control in semiconductor etch: Practice and possibilities. *IEEE T. Semiconduct. M.*, 23(1):87–98, Feb 2010.

J.A. Rossiter and J. Richalet. Realigned models for prediction in MPC: a good thing or not? In *Proc. of the 6th Advanced process control Conf.*, pages 63–70, 2001.

R. L. Stenzel. Microwave resonator probe for localized density measurements in weakly magnetized plasmas. *Rev. Sci. Instrum.*, 47(5):603–607, May. 1976.

P. E. Wellstead and M. B. Zarrop. *Self-tuning systems: Control and Signal Processing.* Wiley, 1991.