IEEE Trans. CPMT (C), Vol. 19, No. 4, pp. 307-314, Oct. 1996.

Run by Run Control of Chemical-Mechanical Polishing

Duane Boning, William Moyne, Taber Smith
Microsystems Technology Laboratories, Massachusetts Institute of Technology,
Cambridge, MA 02139

James Moyne, Roland Telfeyan
University of Michigan, Center for Display Technology and Manufacturing, University
of Michigan, Ann Arbor, MI 48109

Arnon Hurwitz, Scott Shellman
Statistical Methods and Advanced Process Control, SEMATECH, Austin, TX 78741

John Taylor
Compugenesis, Inc., Austin, TX 78739

Abstract

A prototype hardware/software system has been developed and applied to the control of single wafer chemical-mechanical polishing (CMP) processes. The control methodology consists of experimental design to build response surface and linearized control models of the process, and the use of feedback control to change recipe parameters (machine settings) on a lot by lot basis. Acceptable regression models for a single wafer polishing tool and process were constructed for average removal rate and nonuniformity which are calculated based on film thickness measurement at nine points on 8” blanket oxide wafers. For control, an exponentially weighted moving average model adaptation strategy was used, coupled to multivariate recipe generation incorporating user weights on the inputs and outputs, bounds on the input ranges, and discrete quantization in the machine settings. We found that this strategy successfully compensated for substantial drift in the uncontrolled tool’s removal rate. It was also found that the equipment model generated during the experimental design was surprisingly robust; the same model was effective across more than one CMP tool, and over a several month period. We believe that the same methodology is applicable to patterned oxide wafers; work is in progress to demonstrate patterned wafer control, to improve the control, communication, and diagnosis components of the system, and to integrate real-time information into the run by run control of the process.

1.0 Introduction

The chemical-mechanical polishing (CMP) process is of critical importance to current and future generation interconnect for integrated circuit technologies. In addition to CMP process and equipment development [1], the modeling of CMP processes is an active area of research, including work on wafer scale dependencies [2], feature scale models [3], as well as behavior of the equipment over many runs [4, 5]. The challenges posed by CMP for both sensor and control research are also becoming better known [6, 7]. While a good deal of research into run by run control methods has been reported (see [8] for a review), relatively little practical experience with CMP control exists [7, 9, 10]. In this paper, we present the application of an integrated hardware/software control system utilizing run by run methods to overcome common CMP process and equipment difficulties.

In Section 2, we briefly review CMP process and equipment fundamentals, and identify the difficulties accommodated through run by run process control. Section 3 presents the overall control system architecture, and describes the gradual mode run by run control strategy. The polynomial response surface and linear control models developed for CMP are discussed in Section 4. In Section 5, we present a pair of simulation and fabrication experiments that demonstrate the importance and effectiveness of model adaptive run by run control. Finally, in Section 6 we draw conclusions based on these experiments, and highlight areas where additional research and demonstration are needed.

2.0 The CMP Process

In the CMP process, the wafer is affixed to a wafer carrier (via back-pressure), and pressed face-down on a rotating platen holding a polishing pad, as illustrated in Fig. 1. A slurry with abrasive material (e.g. silica particles of sizes from 10nm-200μm) held in suspension is dripped onto the rotating platen during polish. The carrier and platen rotate at variable speeds, typically on the order of 30 rpm. Tools differ in the number of wafers that may be simultaneously polished; single-wafer, dual-wafer, and other multi-headed tools exist.

The process removes material at the surface of the wafer through a combination of mechanical and chemical action. A typical process goal is to achieve "global" planarization (across tens of

mm) by preferential removal of “high” material on the wafer. The planarization of dielectric (silicon dioxide) layers between multilevel metallization steps is one common application. Metal planarization is also often performed.

The control of CMP is chronically poor, arising from poor understanding of the process, degradation (wear-out) of polishing pads, inconsistency of the slurry, variation in pad physical properaties, and the lack of in-situ sensors. Because the process includes mechanical abrasion of the surface, the polishing pad wears rapidly. Concurrent or sequential “conditioning” is usually employed whereby the abrasive surface of the pad is restored (either by mechanical damage to the surface or removal of a thin surface layer), but the lifetime of a pad remains on the order of a few hundred wafers. In addition to difficulties achieving a reliable film thickness (because of changing removal rates over time), the within-wafer uniformity of the polish is difficult to achieve and maintain. Differences between polish rates at the center and edge of the wafer (i.e. “bulls-eye patterns”) may arise due to wafer asymmetry (e.g. wafer flat), non-constant relative pad velocity from the edge to the center, non-uniform slurry and by-product transport under the wafer, wafer bowing due to pressure, or machine drift in time of any of these parameters. As a result of these problems, it is conventional practice to use a number of send-ahead or dummy wafers to condition and/or calibrate the tool before or after each lot of wafers. Here we seek to employ run by run process control to address these process and tool issues - to both reduce or eliminate monitor wafer usage, and to maintain the performance of CMP processes.

In this work, the product characteristics of concern are the removal rate (corresponding to a controlled amount of oxide polished during the step) and the within-wafer uniformity of that removal rate across the wafer. The removal rate is determined by measurement of oxide film thickness before and after polish at each of nine sites on the wafer as shown in Fig. 2, divided by the (fixed) polish time. The “removal rate” output is the average of the nine sites on a wafer. The “nonuniformity” output parameter is computed for each wafer as the standard deviation of the amount removed over the nine sites on the wafer, divided by the average amount removed over the nine sites, times 100.

Two sets of control experiments (detailed in Section 5) were conducted, each on a different polisher. For each experiment, conventional slurry compositions, composite pads, and carrier films were used. Pad conditioning was performed simultaneously or sequentially, depending on

the tools used. Film thicknesses were measured on a Prometrix Spectromap SM2000. In both experiments, the control goal is to maintain a target removal rate and within-wafer nonuniformity in the face of equipment drifts (primarily pad related) and disturbances. The change in removal rate and nonuniformity for a typical uncontrolled or baseline oxide polish process (with a fixed recipe) is shown in Fig. 3. This run, and all those described in this paper, were conducted on single wafer polishing tools, using 8" silicon wafers with a thick blanket oxide deposition. Because we are only measuring and compensating for wafer-level variation from run to run, we believe that the same methodology is also applicable to patterned wafers in which similar wafer-level variation is typically observed [11].

3.0 Control System Description

The observed drift in CMP processes, and the availability of post-process measurements, motivate the use of a run by run control strategy. A generic semiconductor control system framework has been under development [12-15], and is applied to the CMP control problem. Components of the control system design include:

- Generic cell control - a framework for the implementation and incorporation of specific modules (e.g. for run by run control, communications, metrology, etc.). Generic services provided by the GCC include user interface, database access, and application module communication.
- SECS-II communication with measurement and fabrication equipment. This enables capture of post-process wafer measurements, and of real-time equipment signals, as well as down-load of modified process recipes.
- Statistical process control - conventional and multivariate charting, and mechanisms to dispatch to rapid or gradual mode controllers based on equipment state.
- Rapid process adjustment - for rapid update of existing recipes (e.g. in response to significant shifts in equipment response), or rapid process evolution and optimization.
- Run by run control - for gradual compensation, on a lot-by-lot basis, of equipment and consumables drift.

In this paper, we focus on a subset of these modules used to demonstrate process control for CMP, as illustrated in Fig. 4. The collection of real-time data and preliminary data compression

and diagnosis approaches (e.g. similar to that for plasma etch [16]) will be reported at a future time. A detailed discussion of the Generic Cell Controller (GCC) is available in Moyne et al. [12-15]. A general introduction to run by run control is presented by Sachs [17]; the implementation and extensions of run by run control algorithms used here for CMP control are described in [10], and by Boning et al. [18].

The control architecture of Fig. 4 is expanded in Fig. 5 to highlight the control strategy used in this work. Off-line experiments are performed on the CMP tool to build empirical (static input-output) models of the process response. An optimal process recipe is selected that satisfies (or trades off) design goals; this is used as the initial recipe for process control. Lots of 10 wafers each are planarized in the tool, and measurements of oxide film removal rate and nonuniformity are made on wafers #9 and #10. This information is fed to the gradual mode run by run controller, which adapts the process response models. These updated models are then used to generate a new process recipe which (a) achieves the best (weighted) trade-off among the multiple output targets, or (b) achieves all targets with the smallest (weighted) change in the recipe. The revised recipe is then used for the next lot of wafers.

4.0 Control Model Development

Screening experiments with a set of seven parameters were initially performed to determine which machine parameters exert the strongest influence on removal rate and nonuniformity. Based on this initial screening, an experimental design for control purposes was restricted to speed, pressure, force, and profile. A central composite design in these parameters was performed, with ranges as summarized in Table 1.

Table 1. Experimental Design Parameters

Factor	Lower Bound	Upper Bound
speed (rpm)	20	40
pressure (psi)	0	7
force (lb)	8	10
profile	-0.9	0.9

Second order polynomial regression models were constructed for removal rate and nonuniformity. Model fits with adjusted R^2 of 0.897 and 0.769 for removal rate and nonuniformity, respectively, were achieved. The response surfaces (for nominal values of pressure and profile) are shown in Fig. 6.

Each polynomial regression model is linearized around the operating point to generate a multivariate model for the gradual mode run by run controller:

$$y = Ax + c$$

where y is the output vector (removal rate and nonuniformity), x is the input recipe vector (speed, force, pressure, and pad profile), A is a 4x2 matrix of model coefficients, and c is a vector of offset terms. In this controller, we only adapt or update the offset terms c, while the gain coefficients A remain fixed. As seen in Fig. 6, the full response surfaces are nearly linear over the entire operating space. More importantly, the surfaces are locally very well-behaved (nearly linear, monotonic) thus justifying a linearization of the models for control. The robustness of these models in the face of equipment change is discussed in Section 5.3.

5.0 Controller Simulations and Experiments

An important aspect of the run by run controller is adaptiveness of the process model. This is illustrated by comparison of two sets of control experiments: in the first, we use a "partial" model update, and in the second we engage the full adaptive controller.

For partial model update, the linear model is updated based on the most recent run, but then combined with the original linear model before generating a new recipe:

$$c_t = \quad (y_t - Ax_t) + (1 - \quad)c_0$$

where the subscript t indicates the (lot) run number, and y_t is the actual output resulting from the recipe x_t. The variable c_0 is the offset term in the original linear model. In this case, a weight coefficient of 0.5 was used for removal rate, and 0.3 for nonuniformity. Under this strategy, the system functions similar to a discrete proportional controller (without integral action).

In the case of full model update, recursive adaptation occurs by an exponentially weighted moving average (EWMA) update of the offset term, again based on the error between model prediction and measurement:

$$c_t = \quad (y_t - Ax_t) + (1 - \quad)c_{t-1}$$

Here c_{t-1} is the offset term used on the previous run. The selection of the weight coefficient   is based on consideration of both noise and sampling (i.e., we measure and average only wafers 9 and 10 out of each 10 wafer lot). Again, weight coefficients of   =0.5 for removal rate and 0.3 for nonuniformity were used.

Simulations were performed for these two cases, as shown in Figs. 7 and 8. An equipment simulation model was assumed with gain coefficients 5% greater or smaller than in the control model, linear drift on the removal rate and nonuniformity, and additive gaussian noise similar to those observed on the baseline run. We see that the partial model update does not adequately compensate for the drift in removal rate, although it does reduce the impact of the drift compared to an uncontrolled process. The full model update, on the other hand, succeeds in removing the drift, with a small bias error due to drift trend and model mismatch as expected [19].

The recipe (also called the machine settings, inputs, or control parameters) suggested by the controller are shown in Fig. 8 for the full model update simulation. We see that full multivariate control action is taking place: all four input parameters are used by the controller to keep removal rate and uniformity on target. In these simulations and experiments, a low nonuniformity target (on the order of 100-200) below that typically achievable was used to keep the non-uniformity below some acceptable value (a future desirable enhancement to the controller is the inclusion of constraints on outputs in addition to or instead of specific target values). Note that in these simulations we assume very fine resolution on input parameters. In the experiments that follow, we incorporate extensions in the algorithm that explicitly account for finite input resolution (or quantization), observable as discrete jumps in the input parameters of Figs. 10 and 12. The algorithms used, including the utilization of input and output weights and constraints as well as input quantization, are described in detail in [18].

5.1 Experiment #1: Partial Model Update

The results of the first control experiment, with only partial model update, are summarized in Figs. 9 and 10. Examining the output result for removal rate, we note that the original recipe deviates significantly from the target (1800 Å/min); this model error is fatal without permanent model adaptation. On each run, the controller generates a new recipe which attempts to fully compensate

for the difference between desired and observed removal rate. This change initially (first 50 wafers or so) moves us closer to the target, but once substantial drift accumulates, the controller has no means to permanently improve the control -- on each new run, the controller generates a new recipe which is an improvement over the *original* recipe, but not an improvement over the *previous* recipe. The controller does not take advantage of the full range of input parameters (they have not hit bound constraints); rather, the problem is that no permanent adaptation or learning is achieved, and the resulting control is poor.

5.2 Full Model Update Experiment

In the second experiment, full (permanent) model adaptation was used. Here we see (Fig. 11) that the drift is compensated much more effectively. Examining the control parameter moves in Fig. 12, we find that the algorithm responds well, producing successively more aggressive control effort to compensate for accumulating drift. At the same time as speed is increased to maintain removal rate, however, we also see that force is decreased to improve uniformity. The trade-off between these two goals (seen in the response surface of Fig. 6) is also apparent in the resulting control action.

We can compare the baseline, partial model update, and full model update experiments in terms of mean square error (mean deviation from target squared) in the removal rate. We find that the baseline gives MSE of 3.2×10^4, partial model update of 129, and full model update 36. We thus see that the full model update strategy performs substantially better than no control or the partial model update strategy.

5.3 Control Model Robustness

An interesting issue is the robustness and stability of the regression linearization models used for control [19]. The original regression model and its linearization (Section 4) formed the basis for the first set of control experiments. These control models were in fact developed on a *different* tool than that used for the control experiments, and were developed several months earlier. We found that the CMP control models appear to capture fundamental aspects of CMP (e.g. removal rate increases with increased polish speed and force, as expected). The error in model gain coefficients (e.g. the gain with respect to polish speed) does, however, contribute to the bias error

observed in controlled outputs. Further work to address this bias is underway (i.e. predictor-corrector approaches to attack persistent trends [20], internal model control for better convergence [21]).

To more directly assess the validity of the control models across time and tool type, upon completion of the first control run, a small experimental design was again performed, this time directly on the tool used for the control experiment. A 40 wafer Box-Behnken design in the four control parameters was performed. The resulting models were found to be surprisingly similar to the original models; some coefficients (e.g. speed term in the removal rate model) were numerically very similar, and the remaining coefficients were all of the same sign as the control models. These modeling and control results are encouraging, and indicate that expensive recalibration of control models may not be necessary.

6.0 Conclusions and Future Work

In this paper, we have demonstrated the successful application of run by run control to chemical-mechanical polishing. CMP processes pose significant problems: lack of endpoint or other in-situ sensors, poor (primarily empirical) process understanding, and chronic equipment drift. An integrated control system is under development to address these problems that incorporates run by run control algorithms, real-time monitoring to identify equipment problems, and a generic cell controller for equipment communication, data management, and control execution. In this paper, we have focused on run by run control demonstration utilizing post-process measurements, linearized response surface models from design of experiments, and gradual exponentially weighted moving average model adaptation and recipe adjustments to maintain process goals. We have also shown the importance of model adaptation for removal rate maintenance. By incorporating multivariate (MIMO) control, weighted constrained recipe generation, and quantization of inputs, we have demonstrated control of removal rate while maintaining adequate uniformity.

This work highlights several areas for future research. First, real-time monitoring by way of data compression (principal component analysis) in combination with function approximation offers the hope of (1) early detection and diagnosis of machine problems, and (2) correlation with machine state (e.g. drift) that can feed into better control of the process. Second, additional practical issues remain for full automation of the run by run controller, including rapid mode integra-

tion, the optional selection of a dead band (to reduce or trade-off the number of control moves if desired), and the support of output constraints (rather than requiring output targets). Third, the investigation of additional run by run control methods appears fruitful, potentially including classical LQG, neural network, or stochastic dynamic programming approaches that can take full advantage of emerging empirical and physical models of the process (particularly those incorporating process dynamics) [22, 23]. These experiments have been performed on blanket (unpatterned) oxide wafers; control of uniformity on patterned wafers is currently being pursued and will be presented in the future. Finally, the issues of control for metal and damascene CMP processes are also of practical and future interest and should be investigated.

Acknowledgments

This research has been supported in part by SEMATECH, by ARPA under contract NC-001-0014, and by the Semiconductor Research Corporation under contract 94-YC-085. We also thank IPEC/Planar for their assistance in these experiments.

References

[1] M. Fury, "Emerging Developments in CMP for Semiconductor Planarization," *Solid State Tech.*, pp. 47-54, April 1995.

[2] S. Runnels and L. Eyman, "Tribology Analysis of Chemical-Mechanical Polishing," *J. Electrochem. Soc.*, vol. 141, no. 6, pp. 1698-1701, June, 1994.

[3] S. Runnels, "Feature-Scale Fluid-Based Erosion Modeling for Chemical-Mechanical Polishing," *J. Electrochem. Soc.*, vol. 141, no. 7, pp. 1900-1904, July, 1994.

[4] A. Hu, X. Zhang, E. Sachs, and P. Renteln, "Application of Run by Run Controller to the Chemical-Mechanical Planarization Process, Part I," *IEEE Proc. of the 15th Int. Elect. Manuf. Tech. Symp.*, Oct. 1993.

[5] A. Hu, H. Du, S. Wong, P. Renteln, and E. Sachs, "Application of Run by Run Controller to the Chemical-Mechanical Planarization Process, Part II," *IEEE Proc. of the 16th Int. Elect. Manuf. Tech. Symp.*, Oct. 1994.

[6] R. Jairath and L. Markert, "Metrology and Process Control Issues in Chemical Mechanical Polishing," *NIST 1995 Semiconductor Characterization Workshop*, Jan. 1995.

[7] A. Hu, H.-P. Dun, P. Renteln, and E. Sachs, "Sensor Development and Process Control for Chemical-Mechanical Planarization of Multilevel Interconnect Devices," *Electrochem. Soc. Meeting*, June 1995.

[8] E. Del Castillo and A. Hurwitz, "Run to Run Process Control: a Review and Some Extensions," submitted to *J. Qual. Tech.*, Feb. 1994.

[9] A. Altman, "Applying Run by Run Process Control to Chemical-Mechanical Polishing of Sub-Micron VLSI: A Technological and Economic Case Study," S.M. Thesis, MIT EECS, May 1995.

[10] W. Moyne, "Run by Run Control: Interfaces, Implementation, and Integration," S. M. Thesis, MIT EECS, May 1995.

[11] R. R. Divecha, B. E. Stine, D. O. Ouma, D. Boning, J. Chung, O. S. Nakagawa, S.-Y. Oh, and D. L. Hetherington, "Comparison of Oxide Planarization Pattern Dependencies between Two Different CMP Tools Using Statistical Metrology," *VLSI Multilevel Interconnect Conference,* pp. 427-430, Santa Clara, CA, June 1996.

[12] J. Moyne and L. McAfee, Jr., "A Generic Cell Controller for the Automated VLSI Manufacturing Facility," *IEEE Trans. Semi. Manuf.*, vol. 5, no. 2, pp. 77-87, May 1992.

[13] J. Moyne, H. Etemad, and M. Elta, "A Run-to-Run Control Framework for VLSI Manufacturing," *Proc. Microelec. Processes, Sensors, and Controls*, SPIE, vol. 2091, pp. 379-390, 1994.

[14] R. Telfeyan, J. Moyne, A. Hurwitz and J. Taylor, "Demonstration of a Process-Independent Run-to-Run Controller," *Electrochem. Soc. Meeting*, June 1995.

[15] J. Moyne, R. Telfeyan, A. Hurwitz and J. Taylor, "A Process-Independent Run-to-Run Controller and Its Application to Chemical-Mechanical Planarization," *Sixth Annual SEMI/IEEE ASMC*, Boston, Nov. 1995.

[16] D. White, "In-Situ Wafer Uniformity Estimation using Principal Component Analysis Methods," S.M. Thesis, MIT EECS, May 1995.

[17] E. Sachs, A. Hu, and A. Ingolfsson, "Run by Run Process Control: Combining SPC and Feedback Control," *IEEE Trans. Semi. Manuf.*, vol. 8, no. 1, pp. 26-43, Feb. 1995.

[18] D. Boning et al., "Practical Issues in Run by Run Control," *Sixth Annual SEMI/IEEE ASMC*, Boston, Nov. 1995.

[19] A. Ingolfsson and E. Sachs, "Stability and Sensitivity of an EWMA Controller," *J. Qual. Tech.*, vol. 25, no. 4, pp. 271-287, Oct. 1993.

[20] S. W. Butler and J. Stefani, "Application of Predictor Corrector Control to Polysilicon Gate Etching," *Proc. Amer. Control Conf.*, June 1993.

[21] M. Morari and E. Zafiriou, Robust Process Control, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[22] E. Zafiriou, H. Chiou, and R. Adomaitis, "Nonlinear Model Based Run-to-Run Control for Rapid Thermal Processing with Unmeasured Variable Estimation," *Electrochem. Soc. Meeting*, June 1995.

[23] E. Zafirious, R. Adomaitis,and G. Gattu, "An Approach to Run-to-Run Control for Rapid Thermal Processing," *Proc. Amer. Control Conf.*, 1995.

List of Figures

Figure 1. Chemical-Mechanical Polish Tool Configuration

Figure 2. Measurement Sites

Figure 3. Baseline CMP Experiment

Figure 4. Control System Architecture with Implemented Components

Figure 5. CMP Run by Run Control Strategy

Figure 6. Response Surfaces for Removal Rate and Uniformity (as Function of Speed and Force)

Figure 7. Partial and Full Model Update - Output Comparison (Simulation)

Figure 8. Full Model Update - Control Inputs (Simulation)

Figure 9. Partial Model Update - Outputs (Experiment #1)

Figure 10. Partial Model Update - Control Inputs (Experiment #1)

Figure 11. Full Model Update - Outputs (Experiment #2)

Figure 12. Full Model Update - Control Inputs (Experiment #2)

Figure 1. Chemical-Mechanical Polish Tool Configuration

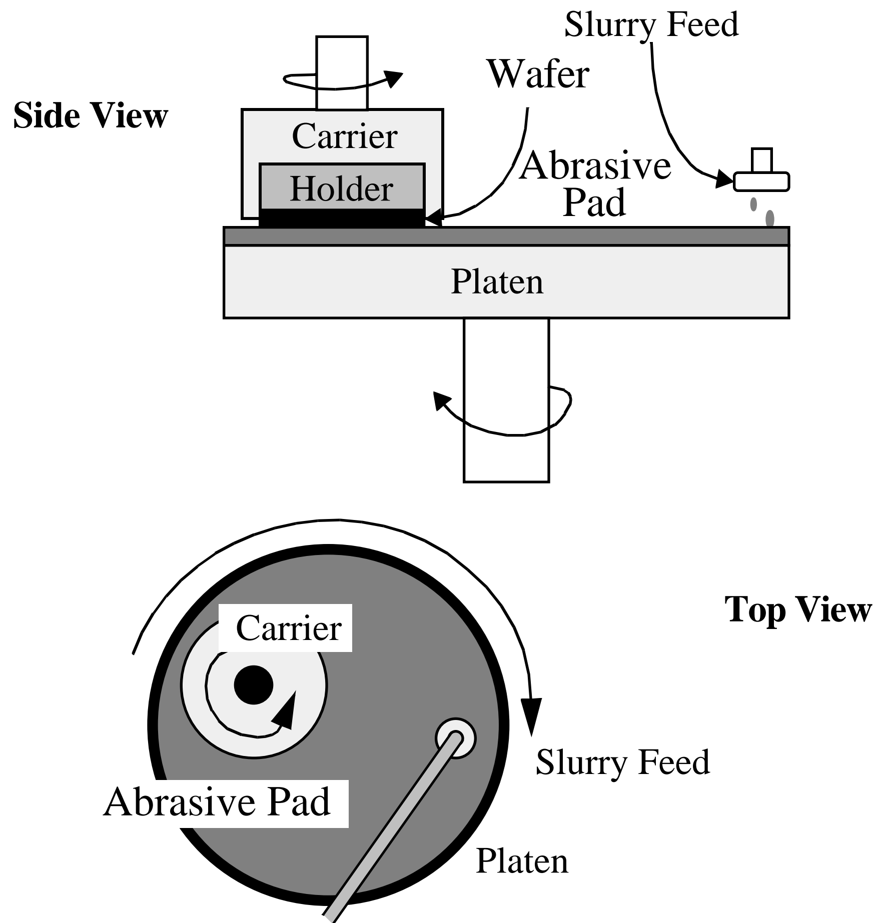


Figure 2. Measurement Sites

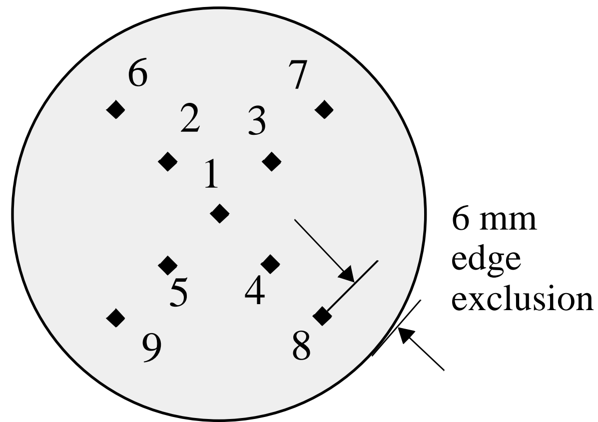
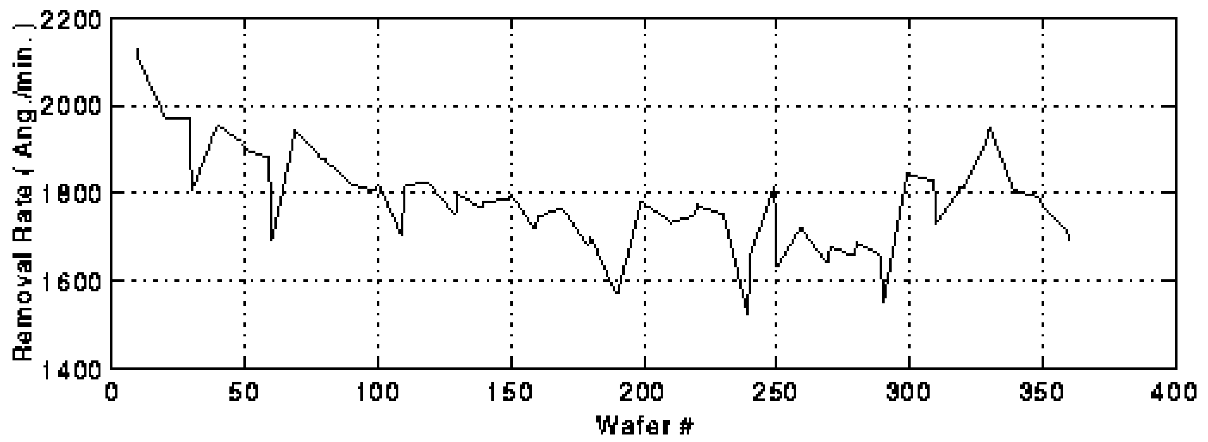


Figure 3. Baseline CMP Experiment



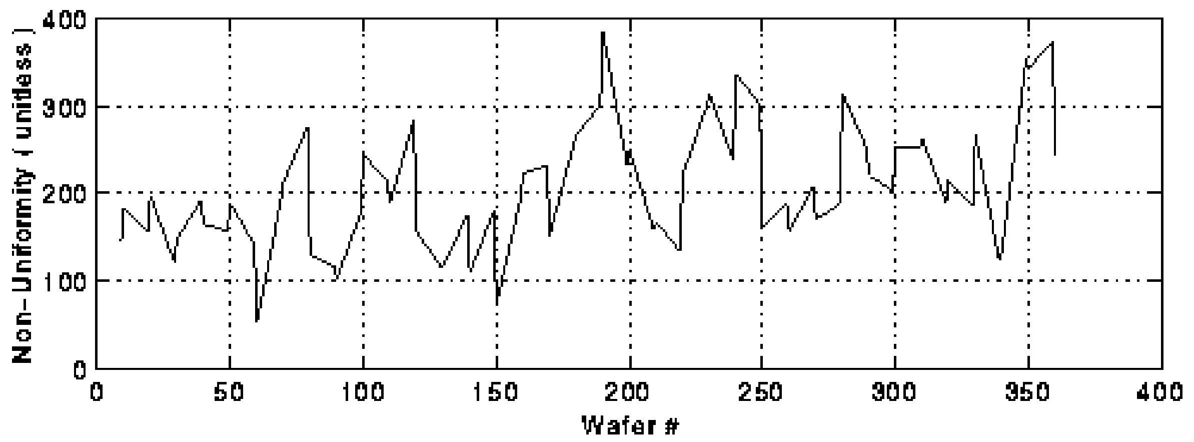


Figure 4. Control System Architecture with Implemented Components

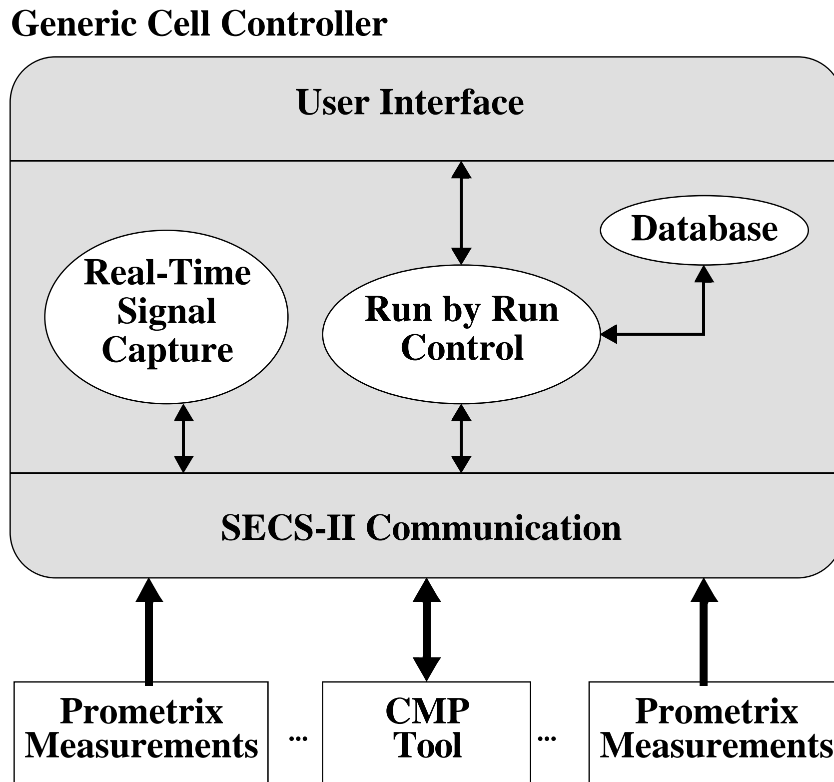


Figure 5. CMP Run by Run Control Strategy

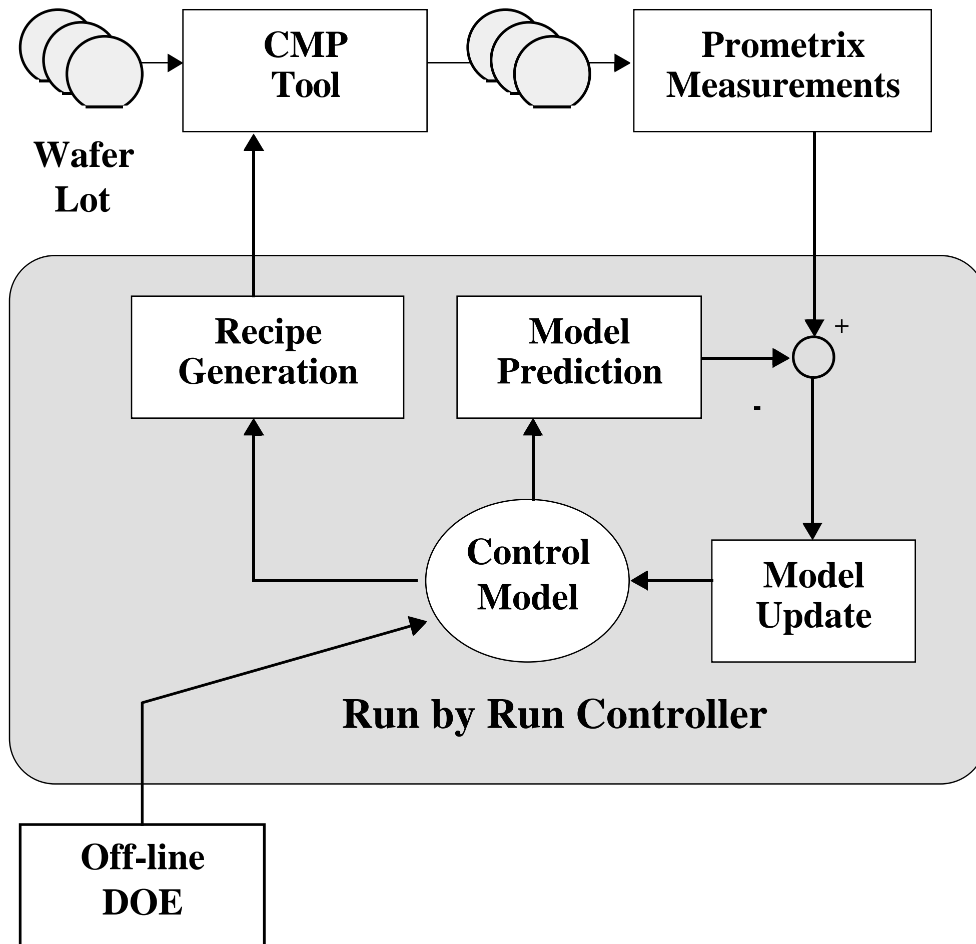
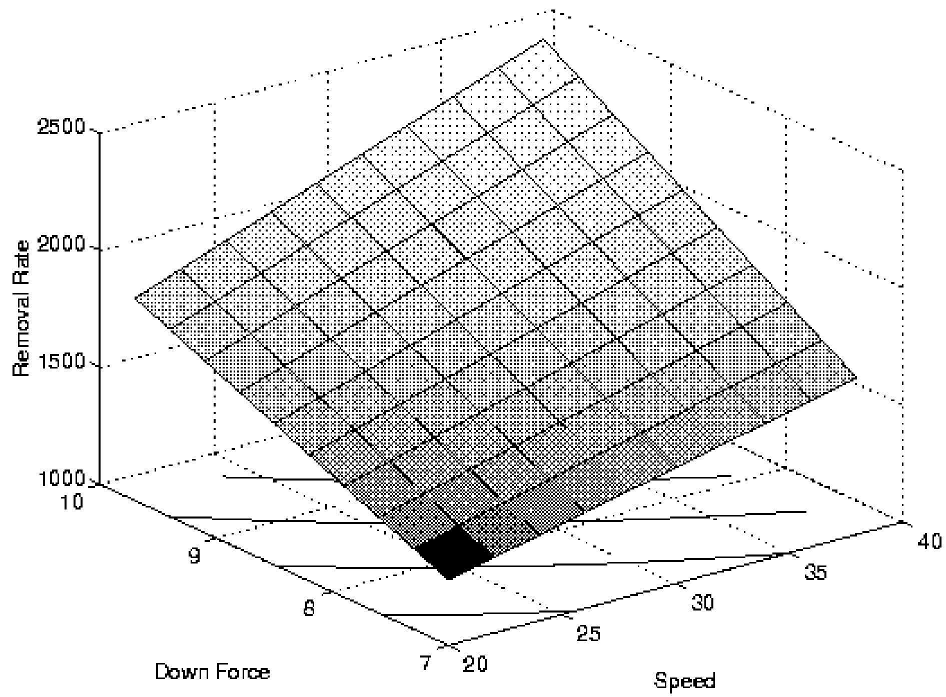


Figure 6. Response Surfaces for Removal Rate and Uniformity

(as Function of Speed and Force)



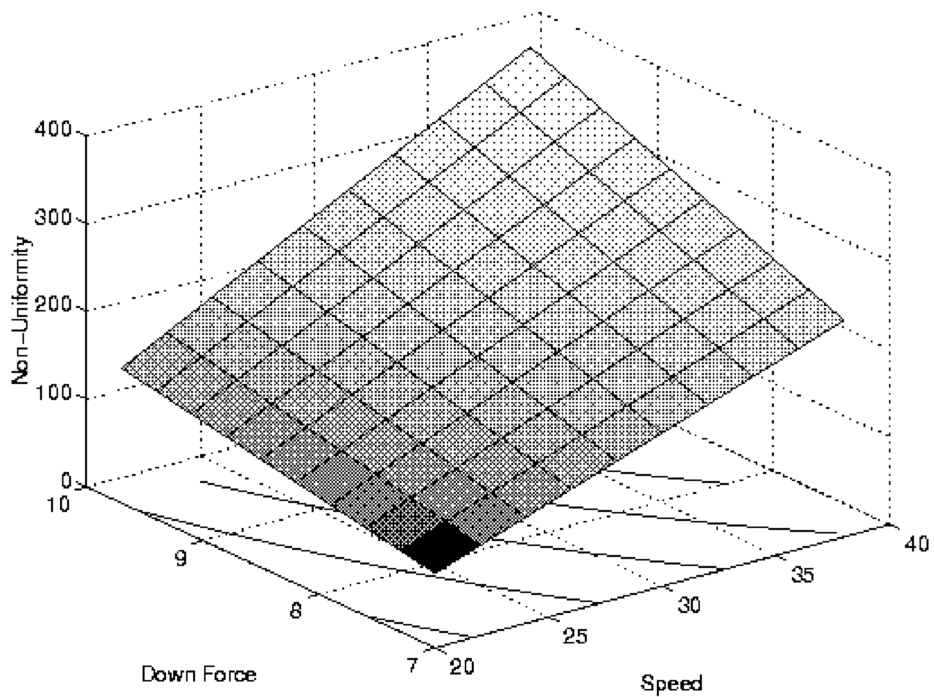


Figure 7. Partial and Full Model Update - Output Comparison (Simulation)

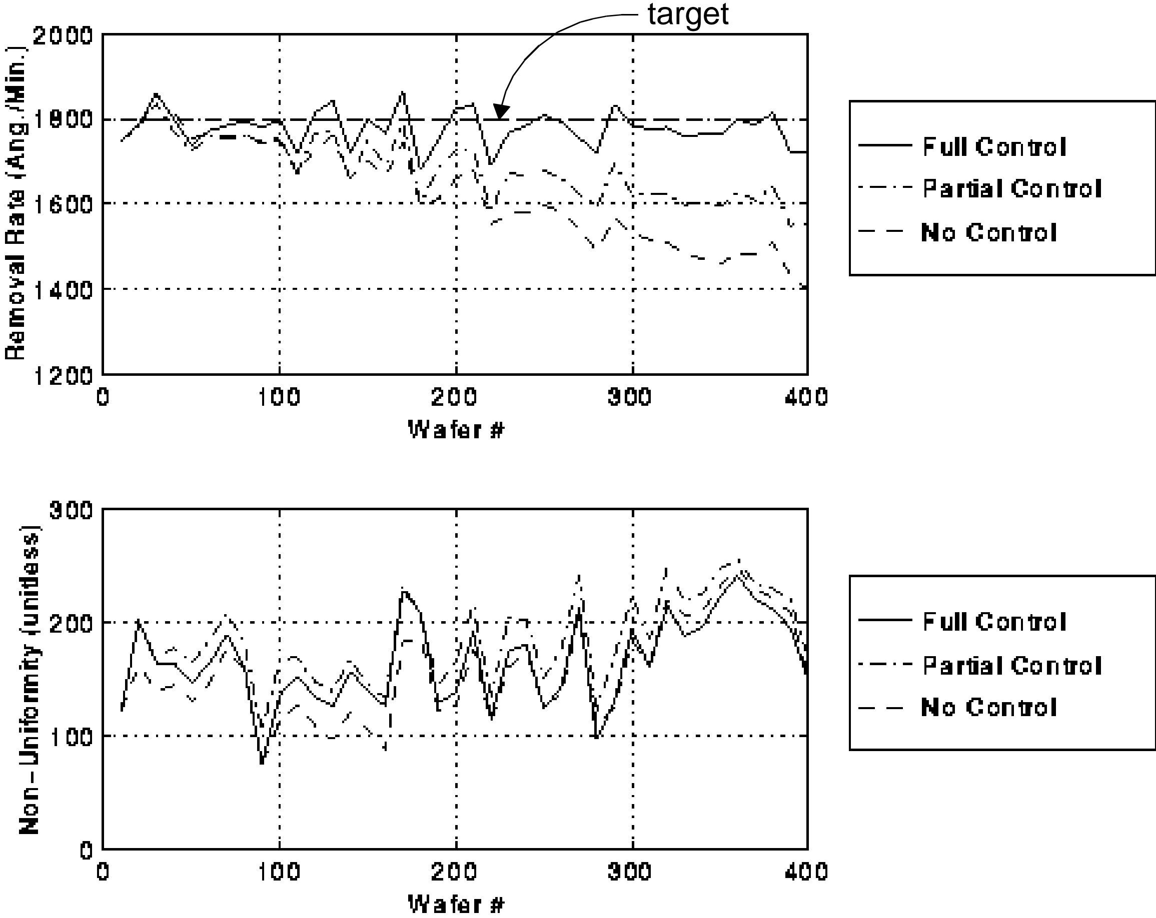


Figure 8. Full Model Update - Control Inputs (Simulation)

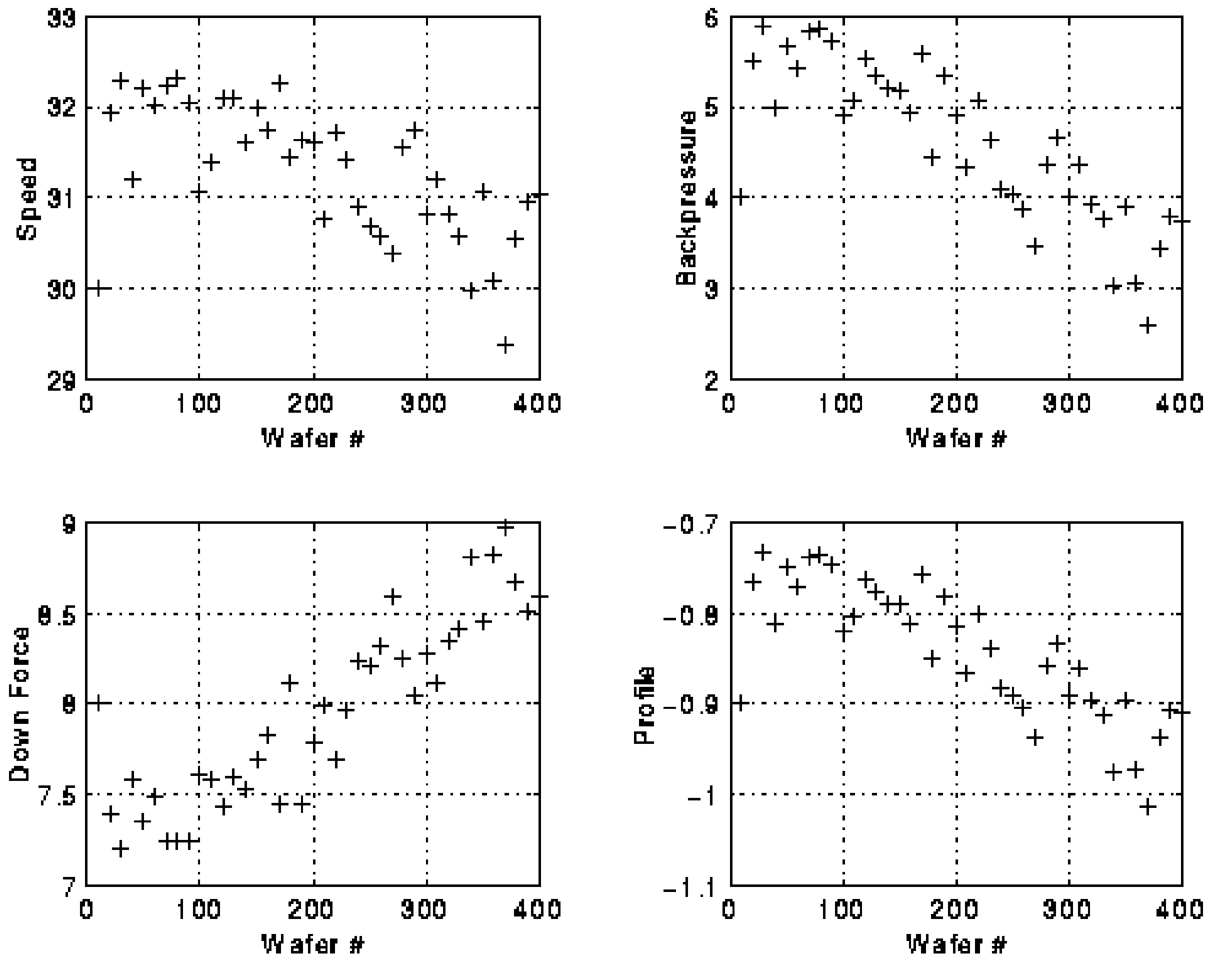


Figure 9. Partial Model Update - Outputs (Experiment #1)

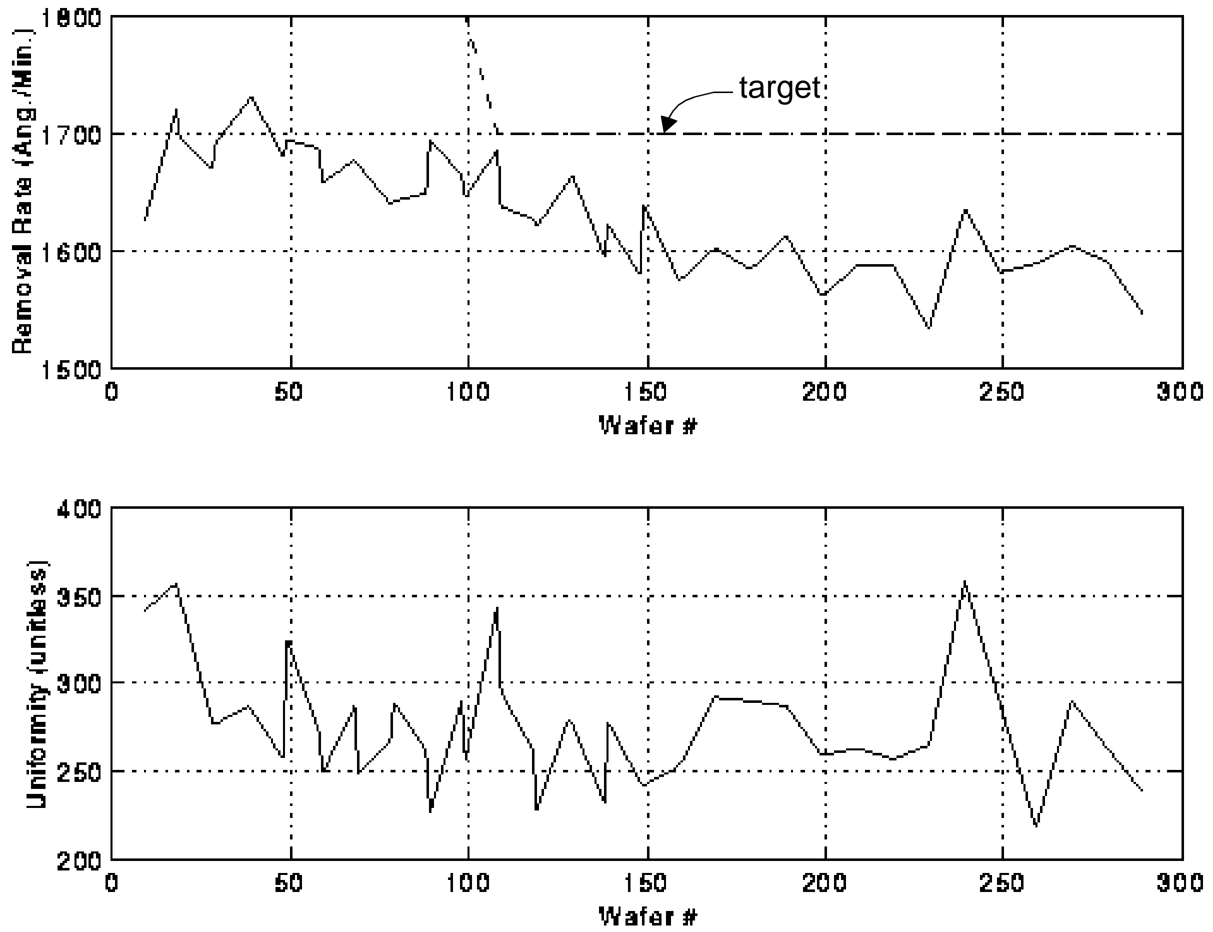


Figure 10. Partial Model Update - Control Inputs (Experiment #1)

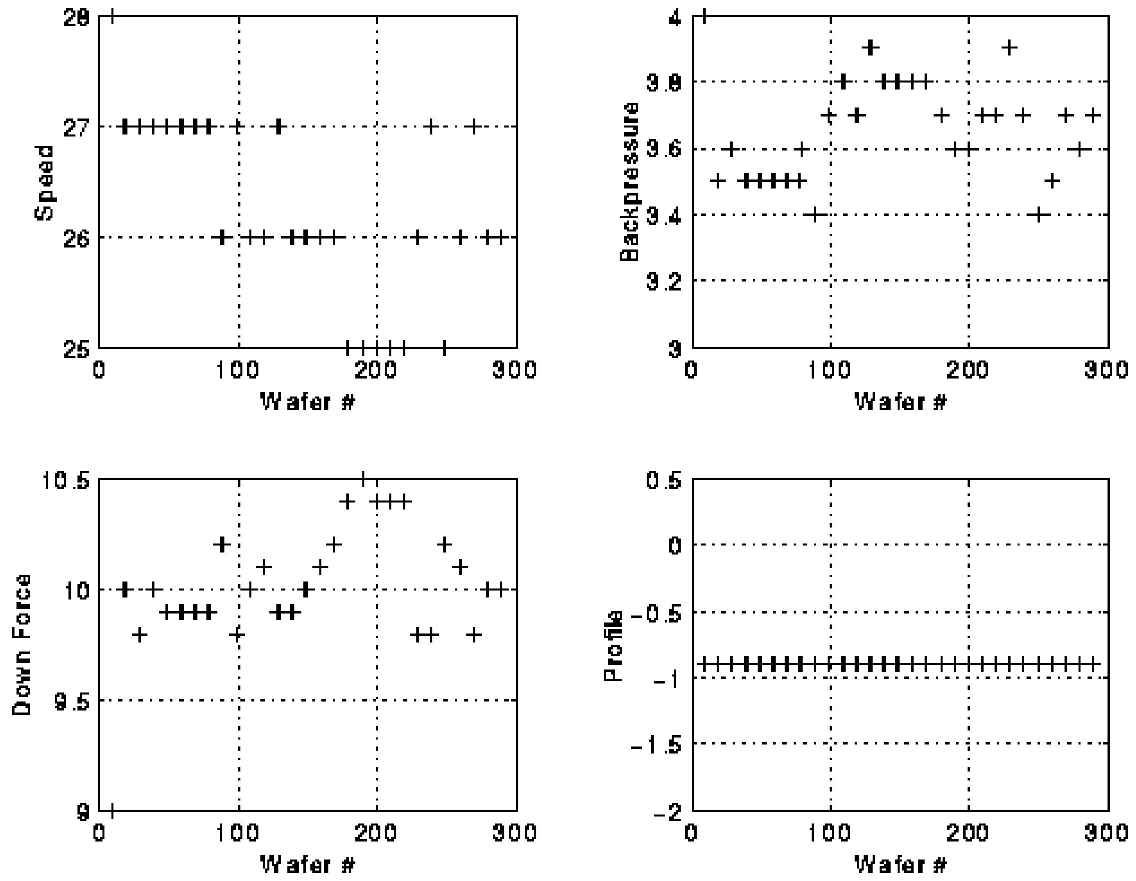
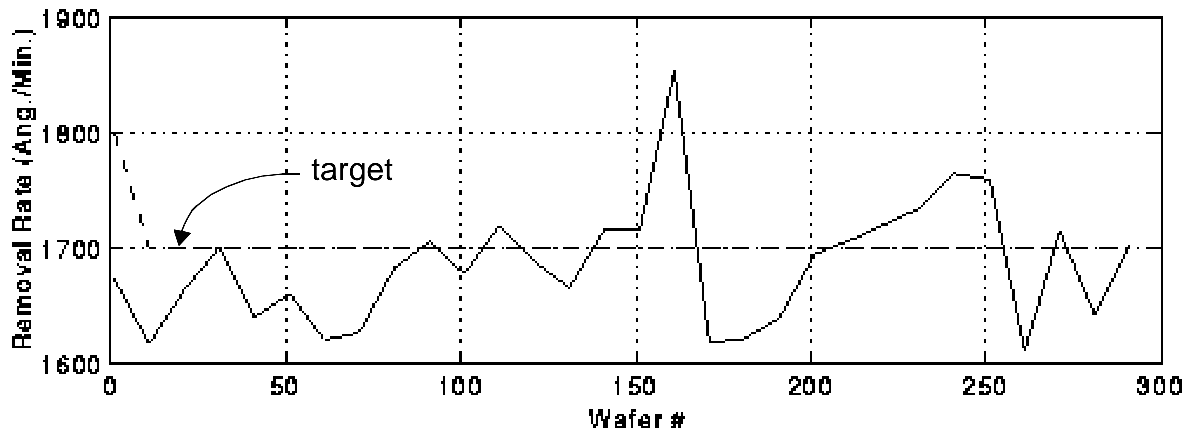


Figure 11. Full Model Update - Outputs (Experiment #2)



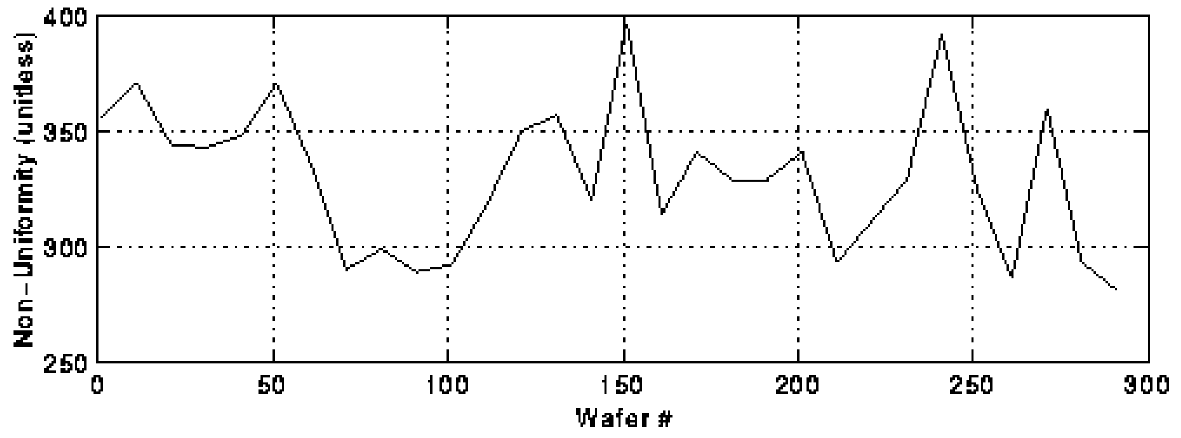
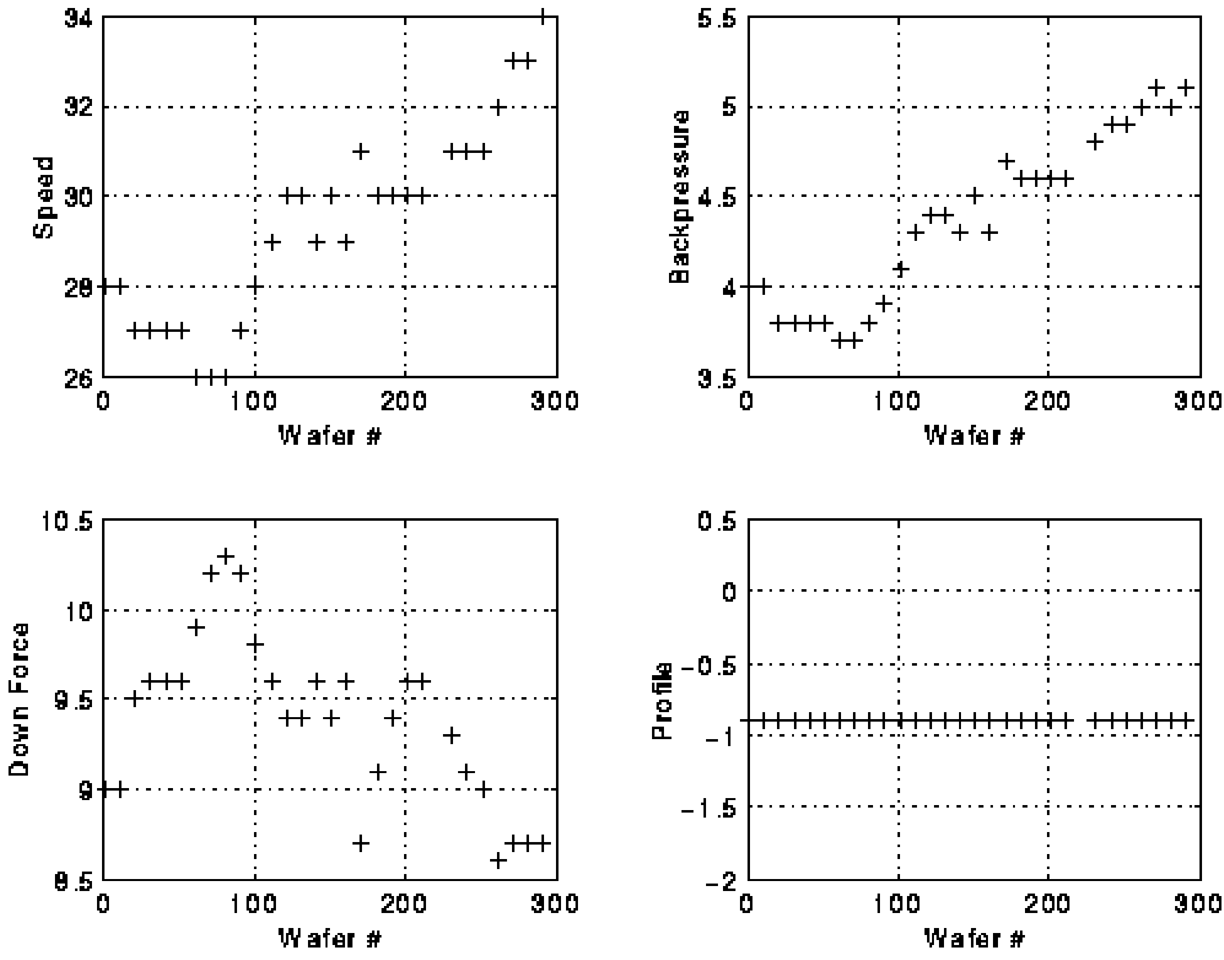


Figure 12. Full Model Update - Control Inputs (Experiment #2)



Author Information

Duane Boning
Room 39-567
Microsystems Technology Laboratories
Massachusetts Institute of Technology
Cambridge, MA 02139
Phone: (617) 253-0931, Fax: (617) 253-9606
Email: boning@mit.edu

William Moyne
Room 36-279
Massachusetts Institute of Technology
Cambridge, MA 02139
Phone: (617) 253-4168, Email: wpmoyne@mit.edu

Taber Smith
Room 39-328
Massachusetts Institute of Technology
Cambridge, MA 02139
Phone: (617) 253-7829, Email: taber@mit.edu

James Moyne
University of Michigan
Display Technology Manuf. Bldg.
2360 Bonisteel, Ann Arbor, MI 48109-2109
Phone: (313) 936-3645, Email: moyne@eecs.umich.edu

Roland Telfeyan
University of Michigan
Display Technology Manuf. Bldg.
2360 Bonisteel, Ann Arbor, MI 48109-2109
Phone: (313) 936-0967, Email: telfeyan@eecs.umich.edu

Arnon Hurwitz
Advanced Process Control
SEMATECH
2706 Montopolis Dr., Austin, TX 78741-6499
Phone: (512) 356-3177, Email: Arnon.Hurwitz@sematech.org

Scott Shellman
Brigham Young University
Email: shellman@statsun.byu.edu

John Taylor
Compugenesis, Inc.
10605 Walpole Lane, Austin, TX 78739-1554
Phone: (512) 288-9376, Email: jtaylor@compugen.com