Model-Based Control for Chemical-Mechanical Planarization (CMP)

Dick de Roover, Abbas Emami-Naeini, and Jon L. Ebert

Abstract— **The research described in this tutorial paper involves an effort for physical modeling and model-based sensing and control of CMP systems. A dynamic model of a rotational CMP process is developed, as well as simulation software. This dynamic model is used for feedback control design based on *in-situ* thickness measurements, as well as run-to-run control using in-line metrology. Simulation results of open-loop, feedback control, and combined feedback and run-to-run control are presented and compared. A multivariable LQ (linear quadratic) feedback controller was designed and showed improvement of Within-Wafer-Non-Uniformity (WIWNU) at the end of a run in simulation over existing open-loop control of a CMP process. It also showed the possibility of using feedback control as a means of end-pointing the CMP process. Furthermore, a run-to-run (R2R) controller was designed and simulated. Additional improvement of WIWNU and tracking a desired average wafer thickness was obtained, showing the merits of a combined feedback / run-to-run control process.**

I. INTRODUCTION AND BACKGROUND

GRINDING and polishing are ancient technologies dating back to antiquity. Modern applications of polishing pertain to glass and lens polishing systems as well as to Chemical Mechanical Planarization (CMP) in semiconductor fabrication plants or "fabs". The technology as applied to semiconductor manufacturing was conceived and developed at IBM in the early 1980s. As in many other innovations, the originators were met with a lot of skepticism. The inventors dared to question the established myths that one could not introduce such a "dirty" technology in a clean room environment, and they paved the way for introduction of CMP as a standard wafer processing step. IBM started ordering custom polishing equipment and the vendors had no idea that their machines were being used in planarizing semiconductor wafers. In 1988, IBM publicly announced its CMP technology. CMP is used to remove material from uneven topography on a wafer surface until a planarized (flat) surface is produced. This enables the subsequent photolithography steps to be done with greater accuracy, and results in film layers to be deposited with minimal height variations across the wafer.

CMP is playing an increasingly critical role in semiconductor microelectronics fabrication [1-3]. The superiority of CMP over traditional etchback techniques with respect to defect reduction and yield enhancement have been demonstrated for tungsten [4]. CMP also has fewer processing steps as compared to traditional etchback methods. And, most importantly, CMP is capable of producing an atomically smooth and damage-free surface which is a basic requirement for semiconductor fabrication below 0.35μ [1]. CMP has been an enabling technology for transition to copper interconnects. CMP is performed in the interconnect structure, where it is repeated in a multiple number of steps. CMP is especially critical to fabricating copper-based semiconductors, where it is used to define the copper wiring structures. CMP has enabled semiconductor device manufacturers to continue shrinking circuits and extend the performance of lithography systems. Optimal CMP will minimize topography, oxide erosion, and dishing in semiconductors.

Integrated Circuit (IC) makers continue to adopt CMP for advanced manufacturing, and CMP has now joined standard processing techniques such as deposition, etch and lithography in strategic importance. State-of-the-art Application-Specific Integrated Circuits (ASIC), and advanced Dynamic Random Access Memory (DRAM) are among the latest applications where CMP is being used. Planarization of topographical features on a semiconductor wafer is a critical factor in ultra large scale integration (ULSI) processing (0.25μ and beyond) for fabrication of multi-levels of wiring and for trench isolation. As device geometries shrink, there are increasingly more stringent requirements on deposition, etch and lithography due to increases in aspect ratio of device structures. There is a lithography constraint on the step height, i.e., topographical variations that require the pattern entirely be confined to within a depth of focus of ± 0.3μ. For DRAM applications, planarization processes for trench isolation require thickness to be controlled within ± 0.1μ or better. This requirement when achieved over all topographies is referred to as "global" planarization. For integrating CMOS technologies of a quarter micron (0.25μ and below), CMP

This work was supported in part by grants from The National Science Foundation (DMI-9983309) and CalTIP.

Dick de Roover, Abbas Emami-Naeini, and Jon L. Ebert are with SC Solutions, Inc., Sunnyvale, CA 94085 USA (408.617.4528; fax: 408.617.4521; e-mail: roover@scsolutions.com).

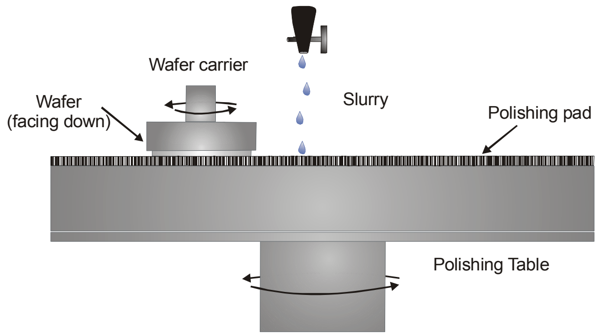

Figure 1. Schematic of a typical rotational CMP System.

is being used in advanced applications such as shallow trench isolation (STI).

A schematic of a typical rotational CMP machine is shown in Figure 1. The rotating wafer rests on a rotating pad system, consisting of two pads, a top pad immediately below the wafer, and a relatively soft sub-pad, located below the top pad. The surface of the pad is not smooth, but contains grooves and is "conditioned" so that the entire surface remains rough. A retaining ring surrounds the wafer and holds it in place. A nominally uniform load pressure distribution acts on the wafer. A uniform pressure distribution of lower magnitude acts on the ring. For oxide or silicon polishing, alkaline slurry of colloidal silica is continuously fed to the pad/wafer interface. To evenly planarize features across the whole wafer, the material removal rate across the wafer must be uniform.

The major problems in CMP are controlling the material removal (or, equivalently, the material removal rate) and the uniformity on each run, and reproducibility from run to run. The goal of CMP processing is to achieve a specified thickness and uniformity in a repeatable fashion. Typically an *in-situ* sensor is used to detect the end-point of the process, i.e., to detect when the desired amount of material is removed, at which point in time the process is stopped. A widely used approach for controlling CMP performance involves the following two-step trial-and-error process: (1) process parameters are tuned to give "good" uniformity, and (2) end-point control using an *in-situ* rate sensor is used to achieve desired material removal thickness. This approach has the following limitations. First, since the process is extremely finely tuned to generate a recipe where the input process parameter values yield acceptable uniformity for most materials, the process operating "window" is very narrow. Therefore, the process performance is not robust, being very sensitive to disturbances and input variations. Furthermore this approach does not work well for several materials. Secondly, if the output specifications for the planarization are changed, then considerable trial-and-error is required to establish the input operating conditions to re-establish uniformity.

An alternative approach that addresses these limitations is a model-based control approach proposed in this paper. We have developed an experimentally validated detailed physical model of the CMP process as well as reduced models for control. We also have developed sensor-based feedback, and combined feedback and run-to-run controller structures applicable to CMP. A software environment, which contains the reduced input-output model of the CMP process as well as the different control strategies, was created to simulate a dynamic CMP clearing process to systematically investigate various combinations of sensing and control strategies for robust control of process performance.

The remainder of this paper is organized as follows. Because of its importance for modeling and control, Section II first discusses sensing. Section III describes the physical modeling of a rotational CMP process. In Section IV control of CMP systems is discussed, including state-of-the-art open-loop control, *in-situ* sensor-based feedback control and *ex-situ* sensor-based run-to-run control, and simulation results comparing these three strategies are shown. Concluding remarks appear in Section V.

II. Sensing for CMP

An exciting opportunity in CMP is to use real-time *in-situ* sensors, such as optical or eddy-current sensors to monitor wafer-scale as well as die-scale uniformity, and allow real-time feedback control of wafer uniformity. The primary objective is to control global planarity. Therefore we have to sense variations (non-uniformity) in the removal rate at the wafer-scale as well as die-scale. Off-line metrology can certainly be used to monitor both wafer as well as die scale uniformity, and the resulting measurements can be used for run-to-run control. It is anticipated that potentially some of the existing optical sensors will be augmented, for example, the Applied Materials ISRM™ sensor, or KLA-Tencor's Precice™ sensor with added pattern recognition capabilities that enable these sensors to monitor wafer-scale as well as die-scale uniformity in real-time. Several sensors have been proposed for and used in CMP for monitoring the material removal rate. Since we are interested in combining run-to-run (R2R) control with feedback and feedforward control, the important sensing issues are:

- the choice of the appropriate sensor for each type of control method;
- the use of *in-situ* optical sensors (ISRM™, Precice™), currently a material removal end-point sensor, monitoring uniformity across the wafer;
- the development of sensor models and estimation algorithms for control;
- validation of the sensor model with dynamics added if necessary.

We will first briefly review the sensing methods currently employed in CMP as shown in Table 1. Two

TABLE 1
SENSORS FOR CMP

Mode	Variable Sensed	Sensor	Control
IN-SITU	Thickness	Optical, Eddy-Current Sensor	R2R, Feedback
	Uniformity	Optical Sensor	R2R, Feedback
	Friction	Motor Torque	R2R, Feedback
	Temperature	Pad Temperature	Feedback
EX-SITU	Thickness	Optical	R2R
	Uniformity	Optical	R2R

sensing approaches are commonly employed in CMP process monitoring: *in-situ* sensors and *ex-situ* sensors. Various sensors have been proposed for *in-situ* measurement, detection and real-time control during CMP processing. Essentially these *in-situ* sensors are of two types. The first type of sensors detects a selected end-point, for example, the point in the process when one layer has been polished and the next layer is exposed (to the polishing surface). The second type of sensor detects *in-situ* process variables such as removal rate and film thickness. These *in-situ* detection (ISD) sensors are generally of two types: Polish Depth (PD) and Remaining Depth (RD) sensors. The former, based on optical reflectance/thin-film interferometry [5], measures the thickness of the material removed, while the latter, based on Michelson interferometry, measures the thickness of the remaining film. An example of an *in-situ* sensor is the *In-situ* Rate Monitor™ (ISRM™) developed by Applied Materials for use in their Mirra™ system. The system's ISRM™ uses laser-interferometry [6] to detect relative changes in the film-thickness during polishing, for highly accurate, real-time process control. The ISRM™ sensor inherently takes material thickness information at several points on the wafer. So, it should be possible to extract Within-Wafer-Non-Uniformity (WIWNU) information from this sensor and, therefore, it should be possible to control WIWNU in real-time.

Another recent *in-situ* sensor for CMP is KLA-Tencor's Precice™ [14]. It uses a combination of optical and eddy current techniques to provide real-time film thickness measurements. The optical system uses a single-wavelength, multi-angle reflectometer. The eddy current probe provides real-time thickness measurement as well. The eddy current probes are inductively coupled resistance measurements to determine sheet resistance in a non-contact manner. Thickness is then deduced from the resistance.

An example of a widely used *ex-situ* sensor is the Nova Integrated Thickness Monitoring System (ITM) [7]. This integrated system can be installed in a location where each product wafer can be measured "wet" immediately after polish within the process equipment. The ITM sensor can be clearly used with a run-to-run controller whereby control variables such as polishing pressure, back pressure, table speed, pad conditioning pattern, and polishing time are adjusted for the next run.

III. MODELING OF CMP PROCESS

A widely used equation for material removal rate in CMP is Preston's equation [11]:

$$\frac{dh(t)}{dt} = K_p p_c(t) v_r(\omega_p, \omega_w), \qquad (1)$$

where $h(t)$ is the material thickness as a function of time, K_p is Preston's coefficient (proportionality constant), $p_c(t)$ is the (possibly time-dependent) interfacial contact pressure, and $v_r(\omega_p, \omega_w)$ is the time averaged relative speed between the point of interest on the wafer and its point of contact on the pad as a function of (possibly time-dependent) platen and wafer rotation speed, respectively. We have developed a 3D contact mechanical model for the contact pressure. This, together with Preston's equation for removal rate and a kinematic model for the relative velocity between wafer and pad, forms the basis of a physical model for CMP. The primary inputs to the CMP process are the rotational speeds of the pad and the wafer (carrier), the load pressure, the ring pressure, and the pad conditioning (modeled as the friction coefficient between pad and wafer). The primary outputs of interest are the uniformity of the material removal rate across the wafer, which can be measured by the Within-Wafer-Non-Uniformity (WIWNU), and the average removal rate.

A. 3D Contact Mechanical Model

We have developed a three dimensional (3D) contact mechanical model, which takes into account frictional force due to the relative motion of the pad with respect to the wafer, to simulate the interfacial pressure between the pad and the wafer. The normalized removal rate distribution obtained from the 3D contact model with normalized friction coefficient of 0.4 compared fairly well to typical experimental data. A coarse model was re-validated by increasing the resolution, hence increasing the number of equations to be solved from approximately 16,000 to approximately 233,000. It was extremely hard to solve this many equations as that would take more than 2 weeks on the older machines. However, currently this system of equations can be solved in approximately 14 hours on a Pentium III, 750MHz with 768MB DRAM.

B. Kinematic Model

Kinematic analysis for a rotational CMP system yields the following expression for relative speed between a point on wafer and pad, see e.g., [12, 13]:

$$v_{rel}(r_w,t)=\sqrt{r_o^2\omega_p^2+r_w^2\left(\omega_w-\omega_p\right)^2+2r_o r_w\omega_p\left(\omega_p-\omega_w\right)\cos\omega_w t}, \quad (2)$$

with r_o the distance between the centers of the pad and the wafer, ω_p and ω_w the rotational speeds of the platen and wafer, respectively, and r_w the distance of a point on the wafer from the wafer center. The time-averaged relative speed of a point on the wafer at distance r_w from the wafer center is now obtained by:

$$\bar{v}_{rel}(r_w)=\frac{1}{T_w}\int_0^{T_w} v_{rel}(r_w,t)dt, \quad (3)$$

with $T_w = 2\pi/\omega_w$ the rotational period of the wafer. For given rotational speeds ω_p and ω_w, and center-to-center distance r_o, the time-averaged relative velocity can be computed as a function of wafer radius by solving the integral (3) numerically. From equations (1)-(3) it is seen that the removal rate is a function of the rotational speeds of pad and wafer, and that perfect uniform removal rate is obtained if and only if pad and wafer rotational speeds are equal, which is a well known fact in CMP processes. If the rotational speeds are not equal to one another, the removal rate will be non-uniform across the wafer. Equation (3) constitutes a static kinematic model for the relative velocity.

C. *Dynamic CMP Model*

As mentioned earlier, we have developed a *static* three-dimensional (3D) contact mechanical model that determines the interfacial contact pressure between wafer and platen, based on specific values for load pressure, ring pressure and average friction coefficient. In addition, a static kinematic relationship was obtained between the platen rotational speed and wafer rotational speed, and the relative velocity between wafer and platen at a given point on the wafer. These static models were combined to compute the static removal rate according to Preston's equation [11]:

$$\frac{dh^i(t)}{dt}=K_p p_c^i(t)v_r^i(\omega_p,\omega_w), \qquad i=0,1,\dots,N \quad (4)$$

where K_p is Preston's coefficient (proportionality constant), $p_c^i(t)$ is the contact (interface) pressure at node i along the radius, and time t, and v_r^i is the time averaged relative speed between the point of interest on the wafer (i) and its point of contact on the pad. The static models were integrated into a dynamic model that predicts the wafer thickness as a function of time-dependent inputs. The dynamic model is shown in Figure 2.

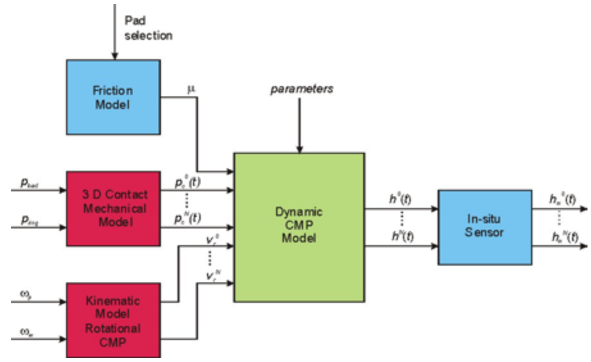

Figure 2. Schematic of dynamic CMP model. $h^i(t)$, i=0,1,...,N represents the wafer thickness in Angstrom for each node i, as a function of time.

In addition to the 3D contact mechanical model (red) and kinematic model (red), a model has also been added for friction, based on pad selection (blue), a CMP *in-situ* sensor model (blue), and a model that computes the CMP dynamics (green) based on customizable model parameters.

We also implemented a model for *in-situ* sensing that can be used for end-point detection algorithms as well as feedback control. Commercially available sensors, such as ISRM™ (Applied Materials) or Precice™ (KLA-Tencor) [14], measure a trace of the wafer thickness along its radius. As our dynamic model has wafer thickness along its radius as state variables, modeling of *in-situ* sensing is done by mapping these state variables to the output of the model. This is the most basic *in-situ* sensing model. A more complicated model could also be derived that takes into account non-perfect measurements by modeling measurement noise, calibration errors, drift, etc.

As most feedback techniques are based on linear models, a linear model has been derived from the non-linear dynamic CMP model that describes the linear CMP behavior in a specific operating point (a selection of constant input values). Comparison of the dynamic behavior between the linear and nonlinear model shows a nearly perfect match in the operating point, and gradual performance degradation when moving away from the operating point. The full-order linear model (201 outputs, 101 state variables and 4 inputs) was further reduced to a 4-output, 4-state, 4-input model for feedback control design. The reduced order linear model was used for feedback controller design using a Linear Quadratic Regulator (LQR) design technique, see Section 4. The dynamic CMP model was implemented in software. Figure 12 shows a full 3D plot of the CMP clearing process with constant values of all four inputs as shown in Table 2.

TABLE 2

NOMINAL INPUT VALUES OF CMP MODEL

Input:	p_l [psi]	p_r [psi]	ω_p [rpm]	ω_w [rpm]
Value:	4.5	2.25	103	97

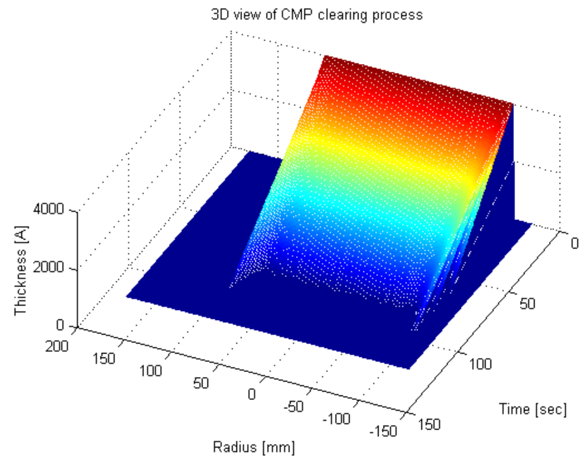

Figure 3. 3D view of CMP clearing process.

Note that the simulation results predict the edge effect that is commonly seen in CMP tools, namely non-uniform clearing at the edge due to a pressure differential at the edge. These results show the potential of the dynamic simulation, both for analysis and feedback control later on.

IV. Model-Based Control for CMP Process

A. Motivation for Control and Controller Structure

The state of the art in CMP control is open-loop control with use of an end-point sensor. As motivated in Section 1, this approach has severe performance limitations. We propose the use of combined feedback control and run-to-run control to achieve optimal CMP performance.

Based on our sensor- and model-based control strategy development, we propose the following feedforward-feedback structure for the CMP controller. The overall control structure is shown in Figure 4 where *u* denotes the control variables (for example, pressure ratio and pad conditioning), *y* the sensed outputs, *z* the performance variables (for example, Within-Wafer-Non-Uniformity (WIWNU), Within-Die-Non-Uniformity (WIDNU) and removal rate), u_{ff} the feedforward control vector, y_{ff} the corresponding reference input vector, and z_{des} the desired planarization performance vector.

The planarization performance, characterized by uniformity (WIWNU, WIDNU) and material removal thickness, are inputs to the input-output model that produce the ideal output, the ideal control and the optimal strategy to deal with process disturbances. There are three major components in the control architecture: the CMP process equipment, the sensors, and the real-time controller. The closed-loop control of the process equipment has three components: the feedback controller or robust regulator, C_{fb}; the feedforward controller (or path-planner), C_{ff}; and the run-to-run controller, C_{r2r}. The ideal operating

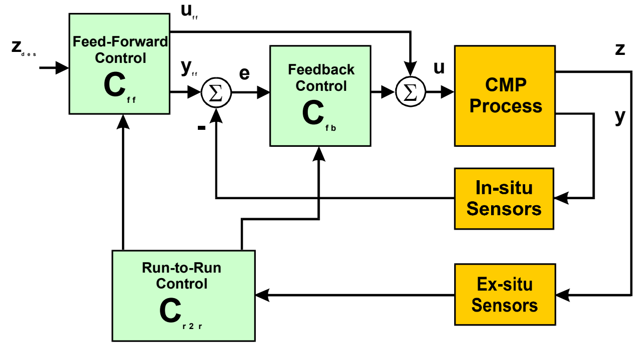

Figure 4. Integrated control architecture for CMP.

setpoints for the CMP equipment are passed on to C_{ff} for a particular set of operating conditions for planarization, e.g., pressure ratio, and pad conditioning. The feedforward controller uses very simple physical models to determine if certain corrective actions will be possible. The feedforward controller will also be responsible for dynamically changing the ideal process conditions such as pressure ratio in order to achieve desired performance. The feedforward controller C_{ff} determines actions based on the desired CMP performance (u_{ff}, the feedforward control input, attempts to minimize $||z\text{-}z_{des}||$). During the process, the estimator, which is part of C_{fb}, will provide estimation of process variables based on models and sensed outputs. C_{ff}, working together with the C_{fb}, will bring the controlled variables to the ideal values when deviations occur so as to minimize the tracking error (u_{fb}, the feedback control signal, minimizes $||y\text{-}y_{ff}||$ despite model uncertainty).

The optimal conditions can be used as set-points for run-to-run control [8]. The run-to-run controller C_{r2r} uses available post-process and/or *in-line* process metrology to adjust control variables from wafer-to-wafer [8]. The run-to-run controller minimizes the effects of drifts and disturbances by adjusting the recipe variables, e.g., the pressure ratio to compensate for pad wear for the purpose of controlling uniformity.

In the following sections simulations of a CMP clearing process are shown for three different control strategies:

1. state-of-the-art open-loop control with constant input pressures (Section B);
2. feedback control of input pressures based on *in-situ* measurements (Section C);
3. feedback control of input pressures combined with run-to-run control (Section D).

The process starts at a wafer thickness of 4000 Å and has to clear to 2000 Å with a desired removal rate of 4000 Å/min, see black line in upper left plot of Figure 5. The performance goals are to obtain a final *average* wafer thickness of 2000 Å and a maximum Within-wafer-non-uniformity (WIWNU) of 3%. The maximum allowed load and ring pressures are 10 psi.

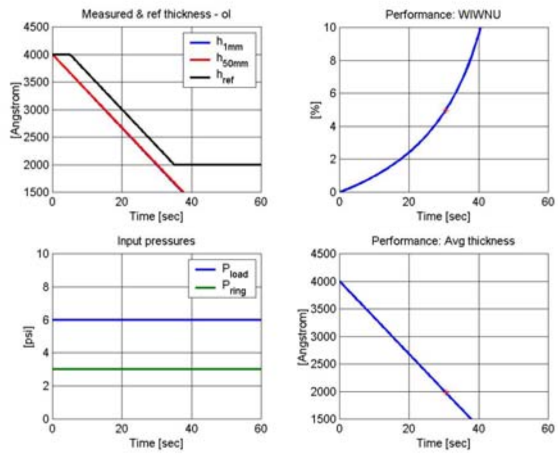

Figure 5. Simulation of CMP clearing process with constant input pressures. Upper left: measured thickness at 1mm and 50mm from wafer center. Lower left: load pressure and ring pressure. Upper right: WIWNU as a function of time. Lower right: Average wafer thickness as a function of time.

B. State-of-the-art Open-loop Control

Figure 5 shows a simulation of state-of-the-art open-loop control of a CMP process with constant input pressures. The values of load and ring pressure are chosen and fixed prior to the process, based on optimization of process results. For this simulation we fixed the load pressure to 6 psi and the ring pressure to 3 psi, as shown in the lower-left plot in Figure 5. Because of its open-loop nature, the process doesn't track the specified reference: it simply clears the CMP process with constant rate of approximately 4000 Å/min. As described in Sections 1 and 2, typically an *in-situ* sensor is used to detect the end-point, which should be 2000 Å in this case. From the upper- and lower-right plots of Figure 5 it is seen that the Within-wafer-non-uniformity (WIWNU) is approximately 5% when the average wafer thickness hits 2000 Å. The WIWNU increases almost exponentially with time as no control is applied.

C. Feedback Control

Figure 6 shows a simulation of a CMP process with feedback controlled load and ring pressure. The feedback controller is a truly multivariable LQ (Linear Quadratic) controller that trades-off performance versus robustness such as non-uniformity, tracking, noise filtering, etc. Note that ring pressure is steadily increased to improve uniformity. From the upper- and lower-left plots of this figure it is seen how the feedback controller tracks the specified reference within a given tolerance. One important aspect to notice is the fact that the feedback controller acts like an end-point detector: as soon as the specified

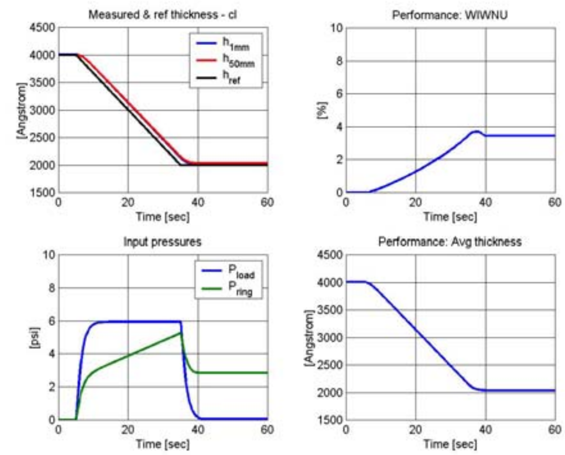

Figure 6. Simulation of CMP clearing process with feedback controlled input pressures. Upper left: measured thickness at 1mm and 50mm from wafer center. Lower left: load pressure and ring pressure. Upper right: WIWNU as a function of time. Lower right: Average wafer thickness as function of time.

reference is achieved, the load pressure goes to zero automatically and ring pressure goes toward a constant steady-state value.

Although this controller has not been optimized, one main advantage of using feedback control is the fact that the controller tries to control the non-uniformity during the run. The final WIWNU for this process is approximately 3.6% (see upper-right plot in Figure 6), with 1.4% improvement in uniformity over the open-loop approach. This also means that if one layer in the stack is not very uniform, the feedback controller will try to maximize the uniformity during the planarization process. In this particular situation, only limited uniformity can be achieved, because there is only one input that affects uniformity, namely the ratio between load and ring pressure. From the lower-left plot it is also seen that the controller increases the ring pressure during the process to try to improve the uniformity.

Another advantage of feedback control, not apparent from this simulation, is the fact that the feedback controller can be used to accommodate disturbances such as process noise and/or drift of machine parameters.

D. Combined Feedback and Run-to-Run Control

The goal of the CMP manufacturing system is to produce multiple wafers, each planarized to a desired WIWNU and desired average removal (rate). WIWNU is typically determined after the product is manufactured, since in most cases sensors are not available which can directly monitor WIWNU during the process. Moreover, during the run, the control variables, or recipe variables, which in our case are the load pressure, ring pressure, are pre-set and held fixed during the run. The run-to-run control problem is to adjust the recipe for the next run based on the results of the previous runs such that the product quality (uniformity, average thickness) improves, i.e., more good product is produced [8]. Run-to-run control has proved to be very popular in process control [9].

The recipe is adjusted from "run-to-run" using the following simple algorithm based on the attributes of the product produced in the previous run or runs [8]. Let $k = 1,2,...,$ denote the run number, $r_k \in \Re^m$ the vector of recipe variables used during run k, $y_k \in \Re^n$ the vector of product quality attributes produced at the end of run k, and e_k the normalized product quality error, whose ith element is defined as,

$$e_k = (y_k(i) - y_{des}(i)) / y_{tol}(i) \qquad i = 1,...,n \qquad (5)$$

where $y_{des}(i)$ is the i[th] desired product quality attribute, and $y_{tol}(i)$ is the associated error tolerance. The error, written in vector form, becomes,

$$e_k = R^{-1}(y_k - y_{des}), \quad R = diag(y_{tol}(1),..., y_{tol}(n)) \quad (6)$$

The simplest choice for run-to-run control is to correct the previous recipe by an amount proportional to the current error. Thus, for run $k = 1,2,...,$ adjust the recipe according to,

$$r_k = r_{nom} + u_k \qquad (7)$$

$$u_k = u_{k-1} - \Gamma e_{k-1} \qquad u_0 = 0 \qquad (8)$$

where r_{nom} is the nominal recipe, u_k is the correction to the nominal recipe for run k, and $\Gamma \in \Re^{mxn}$ is the control design (gain) matrix. It is important to emphasize that (6)-(8) constitute the *complete* run-to-run algorithm. Also (8) has the same form as a gradient descent optimization algorithm. It is possible to choose the run-to-run control gain matrix Γ and to analyze the algorithm under a variety of assumptions about how u_k effects e_k [8]. It can be shown that most of the widely used run-to-run algorithms are in the form of (8) for different choices of Γ.

Figure 7 shows a simulation of a CMP process under feedback and run-to-run control. In addition to the feedback controller described in the previous section, a run-to-run controller of the form (6)-(8) was added that uses an *ex-situ* sensor, i.e., after each run, the measured WIWNU and average wafer thickness are measured and fed into a multivariable run-to-run controller, which updates the load and ring pressures accordingly. The simulation results show the final WIWNU and average wafer thickness after 10 runs. From this figure it can be seen that the combined feedback and run-to-run controller achieves the desired WIWNU of 3%, as well as a desired average wafer thickness of 2000 Å.

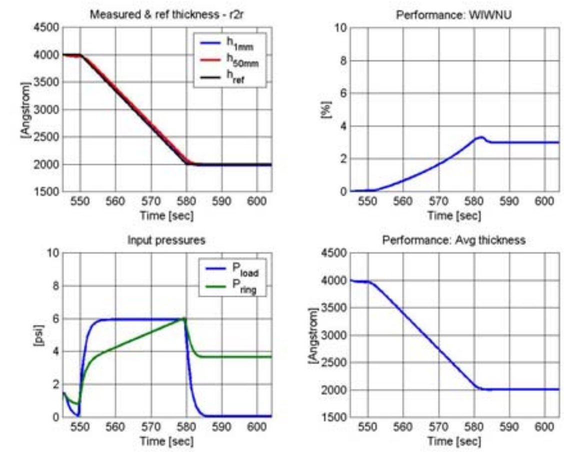

Figure 7. Simulation of CMP clearing process with feedback *and* run-to-run controlled input pressures after 10 runs. Upper left: measured thickness at 1mm and 50mm from wafer center. Lower left: load pressure and ring pressure. Upper right: WIWNU as a function of time. Lower right: Average wafer thickness as function of time.

Figure 8 shows the convergence of the run-to-run controller: the average wafer thickness converged to its desired level after only 4 runs, whereas the WIWNU converged to its desired level after 6 runs. If the run-to-run controller stays on, any future process drifts, such as pad wear, could be accommodated as well.

Finally, Figure 9 shows a comparison of the final radial thickness across the wafer for each of the above-described approaches. The improved uniformity for feedback control and the added improvement of the run-to-run controller are observed.

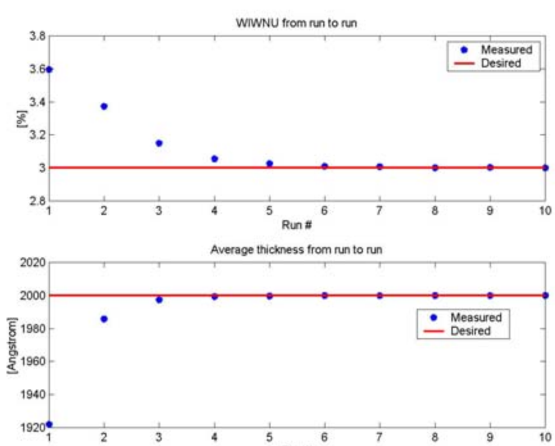

Figure 8. CMP performance from run to run. Upper: Measured WIWNU and desired WIWNU (3%). Lower: Measured average wafer thickness and desired average wafer thickness (2000 Å).

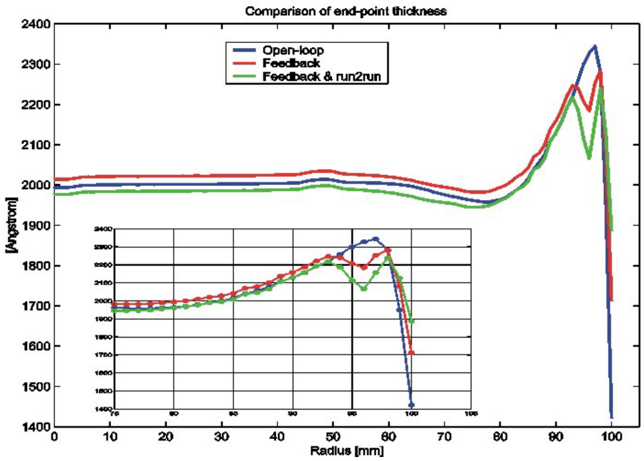

Figure 9. Comparison of end-point thickness for open-loop constant input pressures (blue), feedback controlled input pressures (red) and feedback *with* run-to-run controlled input pressures (green).

V. CONCLUSIONS

This paper discussed recent developments in modeling and model-based control of a CMP process. The following conclusions can be drawn:

A. *Modeling*

- An accurate physical dynamic model for a rotational CMP process has been derived based on a static 3D contact mechanical model, a static model for relative velocity and Preston's equation for removal rate.
- The computed contact pressure compares well to experimental data obtained from a rotational CMP system. Simulations with the dynamic model predict the edge-effect during a typical clearing process.

B. *Control*

- Linear models were derived in several operating points of the CMP process. Based on these linear models, a multivariable LQ (linear quadratic) feedback controller was designed that systematically trades-off performance versus robustness such as tracking, non-uniformity, disturbance suppression, etc. The implemented feedback controller showed improvement of WIWNU at the end of a run in simulation over existing open-loop control of a CMP process. It also showed the possibility of feedback control as a means of end-pointing the process.
- In addition to feedback, a run-to-run controller was designed and implemented in simulation. Additional improvement of WIWNU and tracking a desired average wafer thickness was obtained, showing the power of a combined feedback / run-to-run control scheme.

We will direct future research toward improvement of the proposed control scheme. The feedback control can be improved by directly feeding WIWNU and average wafer thickness measurements into the controller rather than two measurements of wafer thickness. In addition, we plan to expand the model by modeling multiple (>2) pressure points rather than two and designing a controller for this expanded model that will improve the uniformity. Currently, uniformity is limited by the limited system configuration. Additional model enhancements will include the choice of different slurries as well as layer materials.

ACKNOWLEGEMENTS

We gratefully acknowledge the support of NSF and CalTIP through their respective grants.

REFERENCES

[1] The National Technology Roadmap for Semiconductors, Semiconductor Industry Association, San Jose, CA 2003.

[2] J. M. Steigerwald, S. P. Murarka, R. J. Gutman, Chemical Mechanical Planarization of Microelectronic Materials, Wiley Interscience, 1997.

[3] W. J. Patrick, W. L. Guthrie, C. L. Stanley, and P. M. Schiable, "Application of Chemical Mechanical Polishing to the Fabrication of VLSI Circuit Interconnection", J. Electrochem. Soc., 138, 1778-1784, 1991.

[4] K. Wijekoon et al., "Tungsten CMP Process Developed", Solid State Technology, April 1998.

[5] M. Sun, H.-M. Tzeng, H. Litvak, and D. Glenn, "In-situ Detection of Film Thickness Removal During CMP of Oxide and Metal Layers", Proc. CMP-MIC, February 1996.

[6] R. Tolles, T. Osterheld, M. Birang, W. Guthrie, S-H. Ko, T. Cheng, and S. Askari, "Performance of the Applied Materials CMP System for Inter-Level Dielectric Polishing Processes Using Single Step and Multi-Step Process Sequences", Proc. CMP-MIC, pp. 201-208, February 1996.

[7] G. Dishon, et. al., "On-Line Integrated Metrology for CMP Processing," VMIC Specialty Conferences, CMP Planarization, pages 1-5, 1996.

[8] R. Kosut, D. de Roover, A. Emami-Naeini, J. L. Ebert, "Run-to-Run Control of Static Systems", in Proc. 37th IEEE Conf. Decision Control, December 1998.

[9] D. Boning, et. al., "Run by Run control of Chemical-Mechanical Polishing," IEEE/CHMT Int. Elec. Manuf. Symp., 1995.

[10] Timoshenko, S. P., and J. N. Goodier, Theory of Elasticity, Third Edition, McGraw-Hill, 1970.

[11] Preston, F., J. Soc. Glass Technology, 11, 214, 1927.

[12] Yoon, B. U., Young, R. P., In, K. J., Chan, L. S., and Moon,Y. L., "The Effects of Platen and Carrier Rotational Speeds on Within Wafer Non-Uniformity of CMP Removal Rate", 1998 CMP-MIC Conference, IMIC - 300P/98/0193, Feb.19-20, 1998.

[13] Chen, D-Z. and B-S. Lee, "Pattern Planarization Model of Chemical Mechanical Polishing", J. of the Electrochemical Society, 146(2), 744-748, 1999.

[14] KLA-Tencor, Press Release, March 5, 2001.

[15] Handbook of Semiconductor Manufacturing, Y. Nishi and R. Doering, Eds, Marcel Dekker, 2000.

[16] S. Runnels and T. Laursen, "Physics-based CMP Modeling," Solid State Technology, March 2002.