Journal of The Electrochemical Society, **151** (6) G402-G407 (2004)
0013-4651/2004/151(6)/G402/6/$7.00 © The Electrochemical Society, Inc.

Mean Residence Time and Removal Rate Studies in ILD CMP

Ara Philipossian*,z and Erin Mitchell

Department of Chemical and Environmental Engineering, University of Arizona, Tucson, Arizona 85721, USA

Mean residence time (MRT) in the wafer-pad region was shown to be highly dependent on slurry flow rate, wafer pressure, and relative pad-wafer velocity. MRT was also shown to be a linear function of coefficient of friction. The latter was envisioned to be an indication of the tortuosity of the path bounded in the wafer-pad interface. The extent of process transients during chemical mechanical polishing (CMP) was quantified, and it was shown that the average time it took for fresh incoming fluid (*i.e.*, slurry, water, or other active agents) to displace the existing fluid in the pad-wafer region yielded important information regarding fluid concentration near the wafer as well as the kinetics of the process. A new parameter, the turnover ratio, which is defined as the ratio of the MRT to the polish time, was developed to quantify the extent of abrasive concentration transients during a typical polish. This parameter was found to significantly impact the interlayer dielectric (ILD) removal rate and was deemed critical for process optimization considerations.
 [DOI: 10.1149/1.1731539]

Manuscript submitted April 24, 2003; revised manuscript received December 23, 2003. Available electronically May 4, 2004.

One aspect of chemical mechanical polishing (CMP) that requires greater fundamental understanding is the fluid dynamics of the process. Literature states the importance of slurry entrainment in the wafer-pad region;[1-3] however, little has been published on the subject. The foundation of this research is the study of the residence time distribution and mean residence time (MRT) of slurry in the region bounded between the pad and the wafer. The wafer-pad interface can be treated as a closed system reactor, and classical reactor theory can be applied to the slurry flow through the region. To gather critical information about slurry concentration near the wafer surface, experiments were performed to determine slurry MRT (*i.e.*, the average time it takes for fresh incoming slurry to displace an existing fluid in the reactor). MRT is especially critical in processes where multiple fluid steams are required to be introduced to the system sequentially. Because it takes time for a new fluid to replace the existing fluid in the reactor, a period of transient concentration is created which can affect the overall kinetics of the process. Understanding the parameters that have an impact on MRT, and therefore removal rate, is critical to maintain tight specifications in the CMP process, especially as films become thinner and polish times become smaller in accordance with the International Technology Roadmap for Semiconductors (ITRS).[4] By employing classical residence time distribution (RTD) techniques, this study quantifies, in real-time, the extent of slurry concentration gradients for 30 s polish processes and its associated interlayer dielectric (ILD) removal rate as a function of slurry flow rate, wafer pressure, and relative pad-wafer velocity.

Apparatus

A scaled version of a Speedfam-IPEC 472 polisher was used for all tests. The polisher and its associated accessories are described in detail elsewhere.[5] To measure shear force between the pad and the wafer, a sliding table consisting of a bottom plate bolted to the ground and an upper plate bolted to the polisher was used. As contact was made between the wafer and the pad, the upper plate would slide relative to the bottom plate due to friction generated between the pad and the wafer. The degree of sliding was quantified by coupling a load cell to the two plates that would output a voltage to a data acquisition board. The apparatus was calibrated to report the force associated with the particular voltage reading. All polishing parameters were computer-controlled and monitored. In addition, the computer would synchronize the friction table to the polishing process so that real-time friction data, crucial for determining the RTD, could be obtained during polishing. For any given run, the coefficient of friction (COF) was determined by dividing the shear force divided by the normal force applied to the wafer.

* Electrochemical Society Active Member.
z E-mail: ara@engr.arizona.edu

Theory and Experimental

To study MRT at various operating conditions and to examine the dependence of removal rate on MRT, experiments were performed for cases where the wafer was rinsed with water both before and after polishing under dynamic conditions without lifting the wafer from the pad. As such, the polish time included the time it took for the polishing slurry to displace the water under the wafer and for the water to subsequently displace the polishing slurry. During the displacement periods, depending on the magnitude of MRT, the slurry concentration under the wafer would change constantly, thus resulting in a time-dependent removal rate process.

To determine how much ILD is removed during a 30 s polish process using a silica slurry with a solids content of 20 wt %, as presented by curves (a) and (b) in Fig. 1, it is typically assumed that the wafer is subjected to 20% slurry for a prespecified polishing time, t_p, as presented in curve (c). That assumption would hold true only if the slurry replaced the water in the reactor instantaneously. In reality, transient conditions that depend upon the relative magnitude of the polishing time and the slurry MRT must be accounted for in any removal rate model, as demonstrated by curves (d), (e), and (f). For example, curve (f) shows that at high values of MRT, slurry concentration never reaches 20% during the 30 s polishing period. In such cases, the slurry tends to stay in the reactor longer than 30 s. Such long transients influence the amount of ILD removed in two ways. First, because ILD removal rate depends upon the amount of solids in the slurry, the slurry on the wafer surface undergoes extended dilution during the transient period. Second, the presence of extended transients allows the slurry to stay between the wafer and the pad after the slurry source has been shut off, causing the polishing process to continue during the rinsing step. Such complex dependencies necessitate a fundamental understanding of MRT and the factors that influence it during polishing.

This study introduces a new dimensionless parameter that can impact process optimization decisions and removal rates. The parameter, called the turnover ratio (TR), represents the ratio of MRT to the specified polishing time

$$\mathrm{TR} = \frac{\mathrm{MRT}}{t_p} \qquad [1]$$

The experiments consisted of tests aimed at studying the effect of TR on removal rate when the polish includes rinsing steps before and after the polish.

Experiments were performed on Rodel IC-1000 K-groove polyurethane pads. Prior to data acquisition, the pad was conditioned for 30 min using ultrapure water. Conditioning consisted of using a 100 grit diamond disk at a pressure of 0.5 psi, rotational velocity of 30 rpm, and disk sweep frequency of 20 per min. Pad conditioning was followed by a 5 min pad break-in with a silicon dummy wafer.

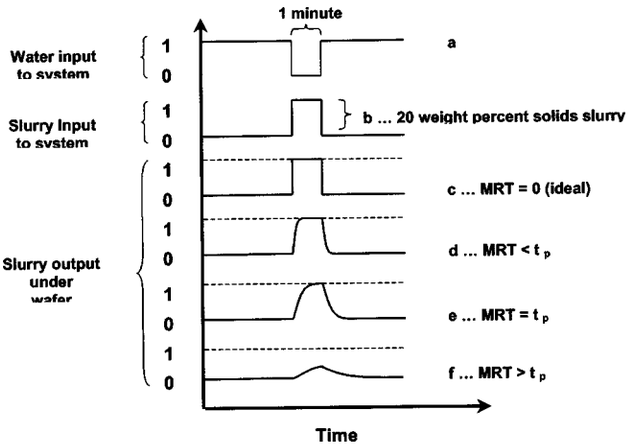

Figure 1. Various scenarios associated with slurry output transients encountered during a 30 s polish process.

Experiments were conducted using *in situ* conditioning at 30 rpm and a disk sweep frequency of 0.33 Hz. Wafer pressures and relative pad-wafer velocities ranged from 2 to 6 psi and 0.31 to 1.24 m/s, respectively. In all cases, slurry flow rate was kept constant at 60 cm^3/min. The Syton-OXK colloidal silica slurry manufactured by DuPont Air Products Nanomaterials was used for all tests. The slurry was selected among various colloidal slurries after extensive tests showed its behavior to be Prestonian[6] over a wide range of operating conditions (Fig. 2).

To calculate MRT, classical reactor design principles were applied[7,8] which state that

$$\mathrm{MRT} = \int_0^\infty t \times E dt \qquad [2]$$

In this case t is time and E signifies the E-curve or residence time distribution of fluid in the system. A new technique was developed

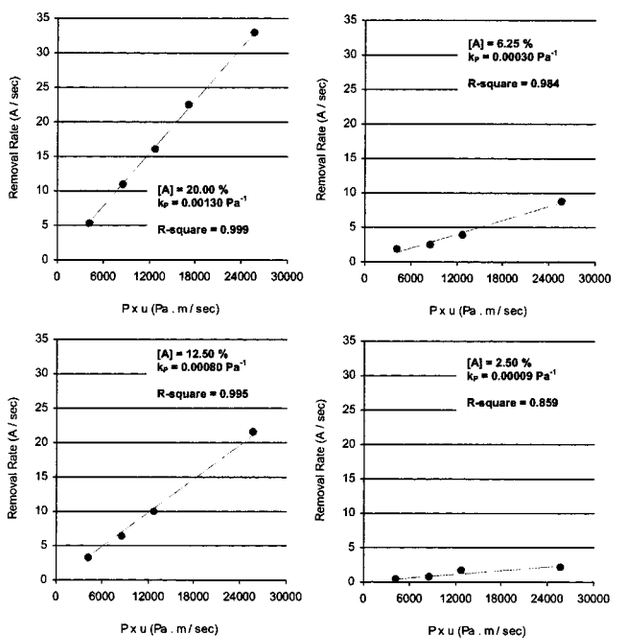

Figure 2. ILD removal rate as a function of $p \times U$ for four solids concentrations (25.00, 12.50, 6.25, and 2.50 wt %). k_p refers to Preston's constant for a given solids concentration.

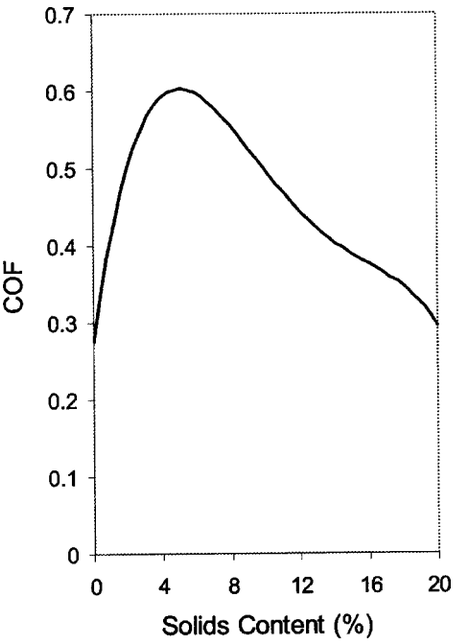

Figure 3. Polynomial fit of experimental data relating silica concentration of the colloidal slurry to COF.

to measure the slurry residence time distribution (E-curve) in the wafer-pad region as a function of slurry flow rate, relative pad-wafer velocity and wafer pressure. This technique relied on the change in shape of the transient response (known as the F-curve) to an instantaneous disturbance within the system (*i.e.*, the sudden replacement of water flow with slurry flow).[5,7-9] The RTD method took advantage of the effect of slurry abrasive concentration on COF to produce and measure a disturbance in the system in order to construct the F-curve, which could then be differentiated to obtain the E-curve.

For each experiment the system was first allowed to reach steady state using ultrapure water as the initial fluid. Ultrapure water was then switched instantaneously to Syton-OXK colloidal silica slurry containing 20 wt% silica. This sudden switch caused the replacement of water in the wafer-pad interface, thus allowing the system to reach a new steady state. Throughout this entire process, COF was measured 1000 times/s. By normalizing the COF response curve (COF *vs.* time), an F-curve was produced from which the E-curve could be constructed. By applying Eq. 1 and 2, MRT and TR could be determined experimentally.

The colloidal silica slurry used for the experiments showed a peculiar relationship between COF and silica abrasives concentration whereby COF reached a maximum as solid content was increased and then dropped substantially at higher concentrations. This is shown in Fig. 3. As a result, the corresponding COF *vs.* time relationship showed an initial increase in COF as the newly introduced slurry began mixing with ultrapure water, followed by a drop in COF as the slurry abrasive content in the pad-wafer region increased. This is shown in Fig. 4. The reasons behind the initial rise in COF as shown in Fig. 3 are not well understood; however, this does not compromise the integrity of the data because the method with which MRT is determined relies on the relationship between COF and abrasive concentration and not on the actual physical or chemical phenomena that may dictate the particular shape of the curve describing the relationship. This nonlinear relationship necessitated the construction of a concentration response in order to produce an F-curve. This was achieved by solving for solids concentration from each COF value on the COF *vs.* time plot using the polynomial fit previously found to correlate the two values. This

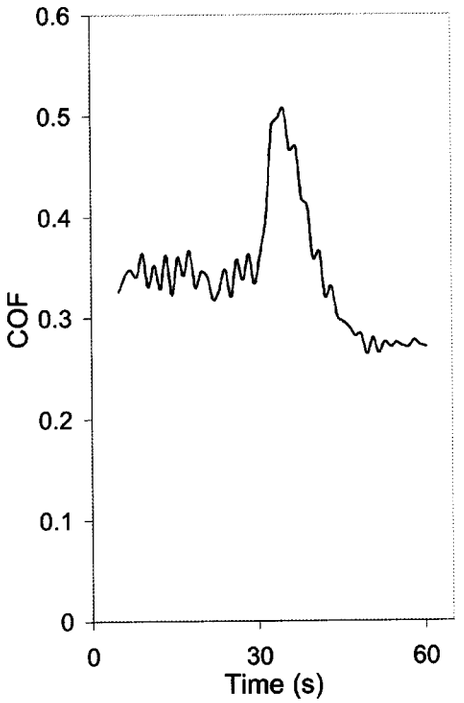

Figure 4. Actual response (COF *vs.* time) corresponding to 20 wt % colloidal slurry displacing water during a polish process.

process yielded a plot of calculated effective slurry abrasive concentration as a function of time. An example of a plot of slurry solids concentration over time, obtained from combining the data contained in Figs. 3 and 4, is shown in Fig. 5. Normalizing and fitting the concentration response would result in the F-curve, which could

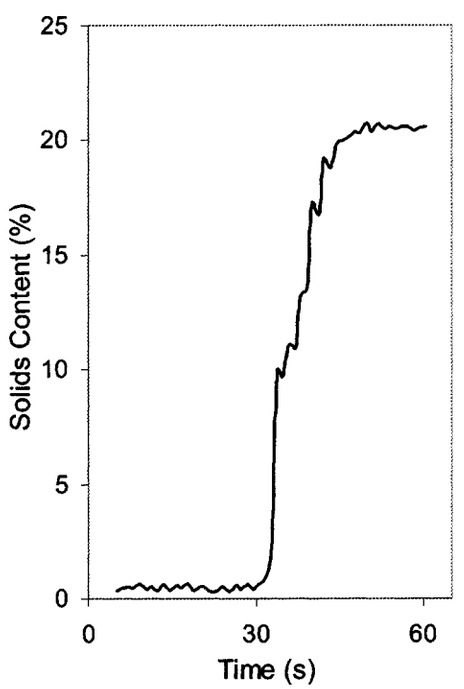

Figure 5. Calculated effective concentration response corresponding to 20 wt % colloidal slurry displacing water during a polish process.

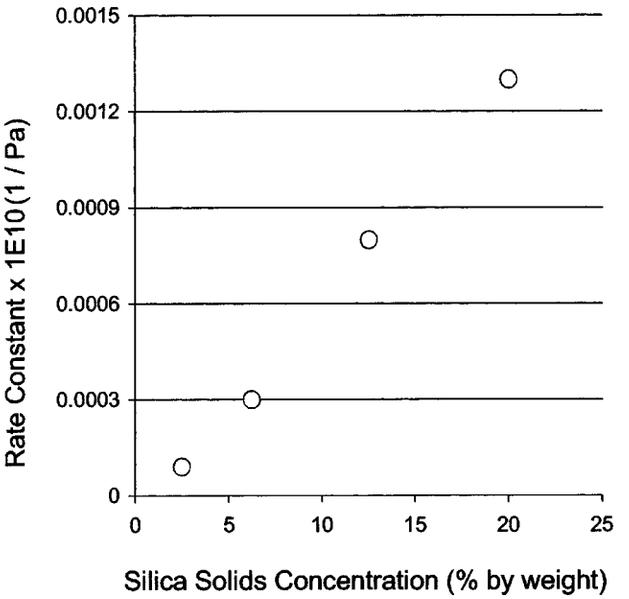

Figure 6. Preston's constant as a function of silica solids concentration in the slurry.

then be differentiated to obtain the E-curve. These procedures are explained in detail elsewhere.[5,7-9]

Once values of MRT were determined for the conditions of interest, the next step was to study how MRT affected ILD removal rate. This was achieved by measuring the removal rate at various TR values. The conditioning and polishing methods were identical to those performed for the MRT experiments. After pad break-in, water was introduced into the system for 30 s, followed by a 30 s pulse of 20% Syton-OXK slurry, which was then followed by a 2 min water rinse step. Because polish time was held constant, TR was allowed to vary with MRT only by selecting different combinations of pressures and velocities. COF was measured 1000 times/s during the entire polish, and its response was converted to a slurry concentration curve using procedures similar to those described previously.

To further validate the results, a technique was developed to model the expected ILD removal from the concentration response. Preston's equation was applied to the plots in Fig. 2 to calculate the rate constant at the four abrasives concentrations and to determine the relationship between the rate constant and silica content. Figure 6 shows the relationship to be linear between silica contents of 0 and 20 wt%. For each test, the concentration response curve was divided into 1 s increments, as shown in Fig. 7, so that the average effective calculated solids content could be graphically determined for each time increment. Using Preston's equation, Eq. 6, and the linear relationship shown in Fig. 6, the expected amount of oxide removed could be calculated for each 1 s time increment. The sum of all time increments would then represent the total expected amount of ILD removed for the entire process. This process was carried out for selected conditions only to help confirm and explain the removal rate data obtained experimentally.

Results and Discussion

The results of MRT as a function of wafer pressure for various pad-wafer velocities are shown in Fig. 8. Each plot also contains COF at that condition averaged over several slurry solids concentrations. Based on at least three repeat experiments, the relative error in measuring MRT was roughly 10%. The average deviation from the mean COF was approximately 20% because the data was reported as average of COF taken at 0, 3, 6, 12, and 20% solids content. At 0.31 m/s, increasing pressure by a factor of three increased MRT and COF by 127 and 48%, respectively. At 0.62 m/s, the increase in

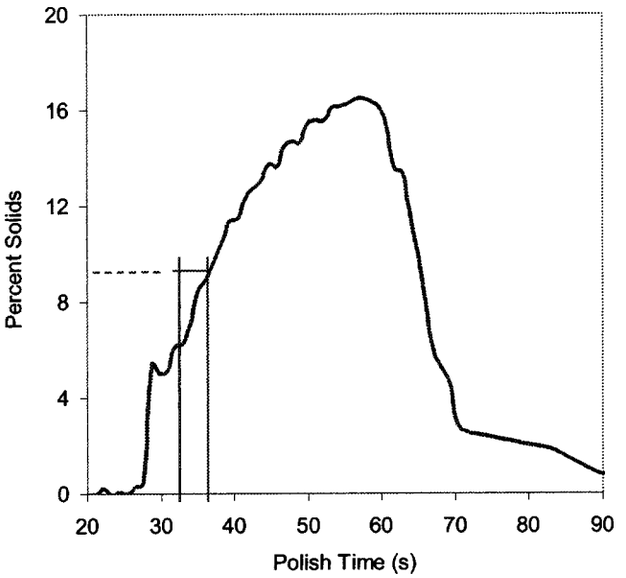

Figure 7. Calculated effective concentration curve showing one time increment used to calculate the expected ILD removal rate.

MRT and COF were 48 and 17%, respectively. At higher velocities, tripling pressure did not affect MRT or COF significantly. Throughout all experiments, the pad and the wafer were in intimate contact with one another,[9] thus indicating a high resistance to flow over the entire polish period. Figure 9 is a plot of MRT as a function of COF averaged over several slurry dilutions. The plot shows an excellent linear fit, suggesting that there is a strong correlation between MRT and COF. This relationship is explained by recognizing that as COF increases, slurry film thickness decreases. Therefore, at low COF values a mostly continuous fluid layer exists between the pad and the wafer with few obstacles to disrupt flow. This leads to low MRT values. With increasing COF values, the fluid thickness continues to decrease and slurry transport mechanisms begin to change as the fluid layer becomes less continuous due to the higher tortuosity of the system (*i.e.*, increased resistance to slurry flow due to the presence of physical barriers such as pad asperities in the wafer-pad interfacial region). In such cases, one can envision pad asperities to act as baffles against the flow, thus increasing MRT.

MRT results revealed that the choice of slurry flow rate, wafer pressure, and relative velocity result in a wide range of MRT. The TR defined in Eq. 1 represents the ratio of MRT to polish time. For this series of experiments, polish time was held constant at 30 s, allowing TR to vary with MRT by selecting different combinations of pressures and velocities. Figure 10 shows calculated effective slurry concentration plots *vs.* time for three different TR values. The

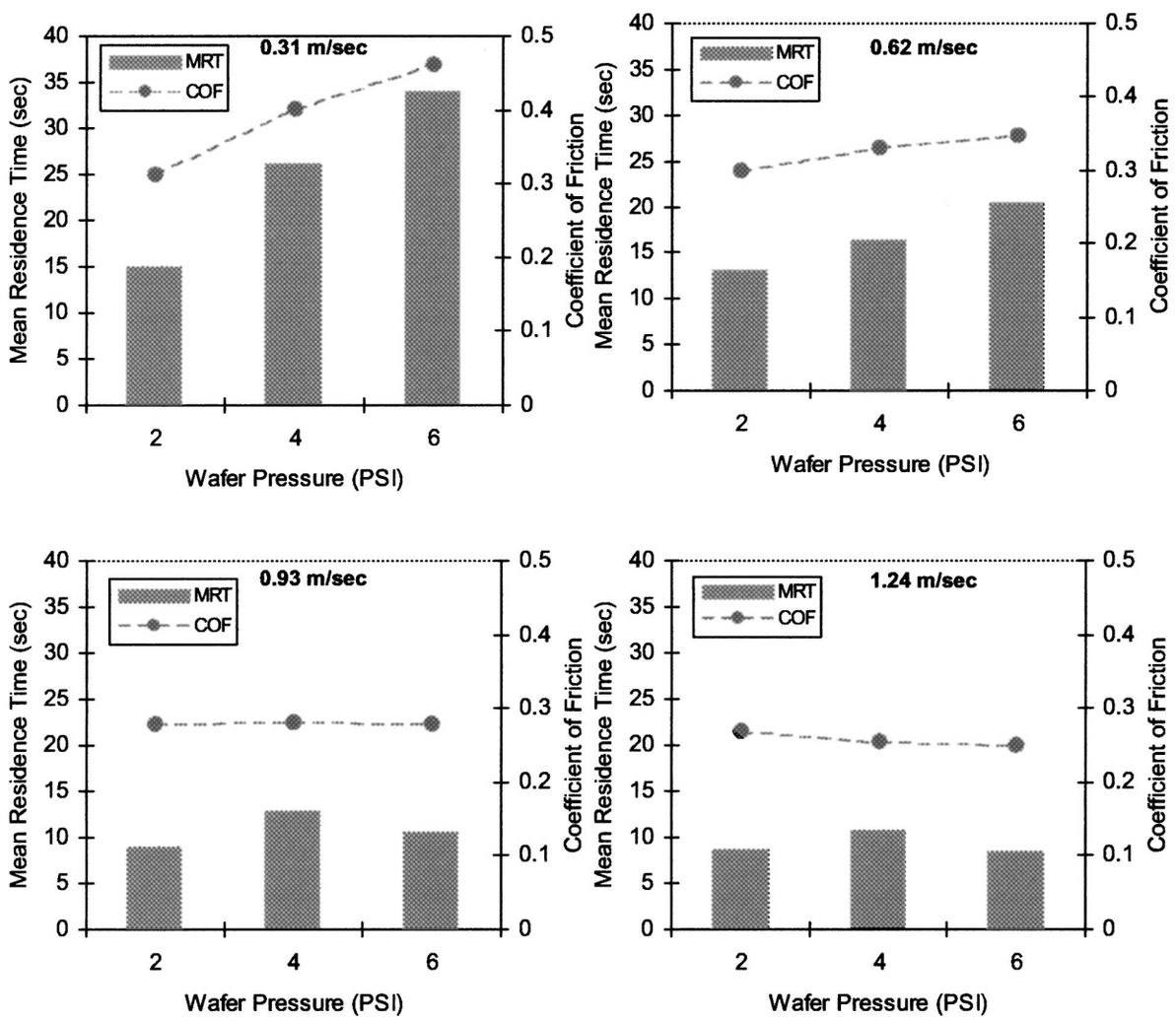

Figure 8. MRT and average COF as a function of wafer pressure for various relative pad-wafer velocities.

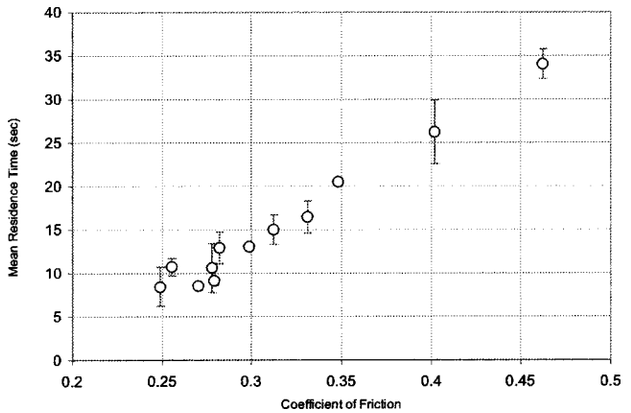

Figure 9. Relationship between MRT and COF.

plots were constructed from COF data using the method described previously. To demonstrate the effect of TR on ILD removal, a method was devised in which MRT could be studied as a function of ILD removal rate independent of the combinations of pressures and velocities required to arrive at different MRT values. This method was realized by using a Prestonian slurry such that different combinations of pressures and velocities could be chosen to alter MRT, as long as the product of $p \times U$ remained constant. In Fig. 10, run no. 1 (*i.e.*, the curve corresponding to the lowest TR) reached a maximum solids concentration of 20% more rapidly than its counterparts. Moreover, the curve corresponding to the highest TR (*i.e.*, run no. 3) never reached the maximum solids concentration during the 30 s polishing period, because the length of the transient period was longer than the polish time. This phenomenon was theorized to negatively impact removal rate because material removal is linearly proportional to silica content as shown in Fig. 6. The trends in Fig. 10 clearly vindicate the theory illustrated in Fig. 1.

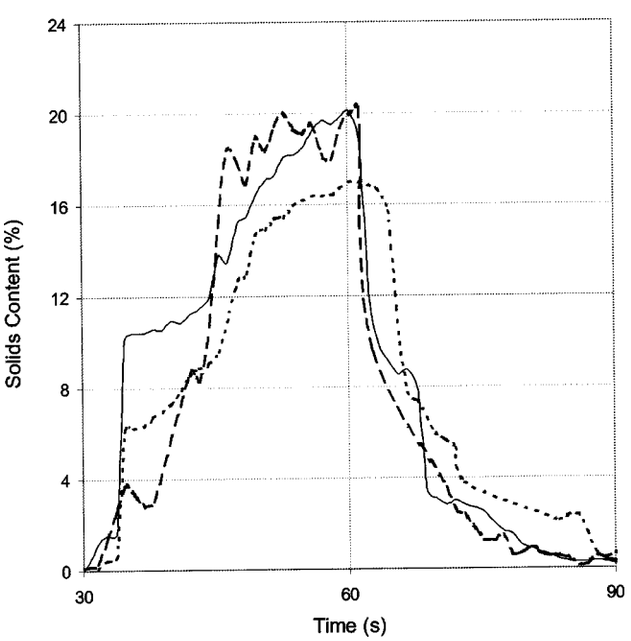

Figure 10. Calculated effective concentration curves corresponding to 30 s polishing processes at three different values of turnover ratio. In all cases, the product of pressure and velocity is 17,099 Pa m/s. (— — —) TR = 0.28; (——) TR = 0.54; (· · · ·) TR = 0.75.

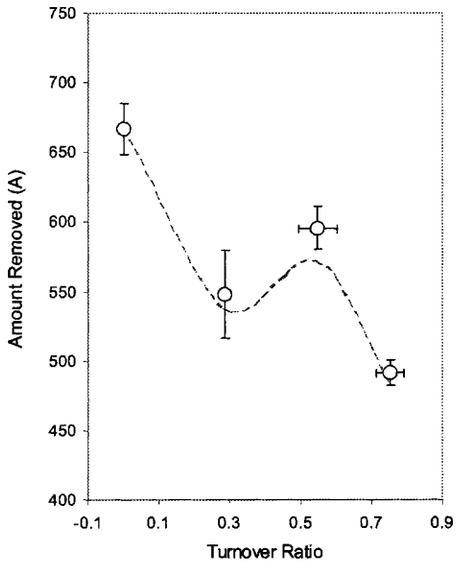

Figure 11. Effect of turnover ratio on ILD removal rate for a 30 s polishing process using a 20 wt % solids colloidal silica slurry: (o) experimental data and (-----) results of the semi-empirical model.

The expected oxide removal associated with each set of conditions shown in Fig. 10 was then quantified from the concentration curves using the method described previously. Figure 11 illustrates the results of the semi-empirical model and experimental data of oxide removal as a function of TR. The data point at a turnover ratio of zero represents the oxide removal when the rinsing steps were eliminated from the polish. The 20% solids slurry was introduced to the system prior to allowing the wafer to contact the pad. This ensured that the slurry solids concentration between the pad and the wafer reached steady state immediately upon polishing. Thus, MRT and therefore TR were equal to zero. The expected oxide removal for this situation was calculated in the same manner as nonzero values, except, because there were no transients, slurry solids concentration was assumed constant at 20% throughout the entire polish. Figure 11 indicates that the semi-empirical model follows a general decreasing trend with TR, as is expected, due to the relatively longer transient periods at high TR. Unexpected, however, is the rise in oxide removal predicted for the turnover ratio of 0.54 corresponding to run no. 2. Referring now to Fig. 10, it is apparent from the concentration curve describing run no. 2 that the CMP reactor experienced a high influx of solids immediately upon replacement of the water source with slurry. Runs no. 1 and 3 do not show nearly as drastic an effect. For approximately the first 10 s of the polish, run no. 2 removed significantly more material than either of its counterparts (up to twice as much as run no. 3 and three times than that of run no. 1). This was the reason that run no. 3 predicted a higher removal than run no. 1 or 2 when the trend suggested that the magnitude of removal should lie between the two. Figure 11 also contains actual experimental data corresponding to the four polishing conditions. The agreement between the data and the semi-empirical model is within 4%, thus indicating the critical role TR played in oxide removal when water was introduced into the system before and after polish. It was also evident that the reason for this was the varying length of transient solids concentration under the wafer depending upon the MRT at the particular condition. In light of these results, additional experimental removal rate data was collected in order to draw a complete picture over a wider range of TR values. ILD removed during the process was normalized to compare removal data taken under different conditions. This was achieved by dividing the total amount removed by the polish time to give the average removal rate. This was then divided by the product of the wafer pressure and relative velocity to yield the average Preston's

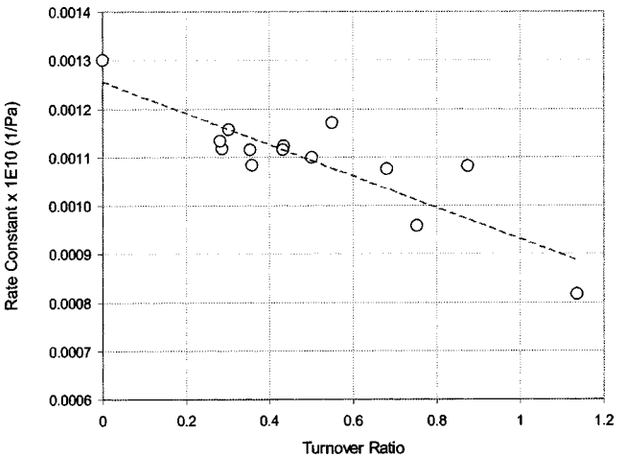

Figure 12. Preston's constant as a function of turnover ratio.

constant, k_p.[6] In such a manner, Preston's constant could be compared across different pressures and velocities because it was independent of those parameters. The operation used to normalize oxide removal is mathematically stated as

$$k_P = \frac{\text{oxide removed } (m)}{30\text{s} \times \text{pressure (Pa)} \times \text{velocity (m/s)}} \quad [3]$$

Figure 12 is a plot of Preston's constant as a function of TR. Each data point is an average of at least three polishes and the relative standard deviation is approximately 10% for both the abscissa and the ordinate. Figure 12 shows that Preston's constant decreases steadily with TR. This gradual drop can be potentially detrimental to polish consistency depending on operating conditions. For example, a 30 s polish operating at 6 psi and a relative pad-wafer velocity of 0.31 m/s has a turnover ratio of 1.13. The linear model predicts that during the 30 s polish, 370 Å of oxide are polished. In the absence of TR effects, the amount removed would be 500 Å. This is a 26% reduction from the expected oxide removal. Operating at pressures higher than 6 psi or velocities lower than 0.31 m/s would reduce the oxide removal by an even greater extent.

Conclusion

MRT in the wafer-pad region was shown to be highly dependent on slurry flow rate, wafer pressure, and relative pad-wafer velocity. MRT was also shown to be a linear function of COF. The latter was envisioned to be an indication of the tortuosity of the path bounded in the wafer-pad interface. COF is much easier to measure directly than MRT; therefore, it would be advantageous to develop a more complete correlation between COF and MRT. This would allow MRT to be predicted by measuring the COF of the system in cases where calculation of MRT is difficult or not practical (*i.e.*, in high-volume manufacturing). Magnitude of the MRT is crucial for single platen processes, because effective displacement of fluid (*i.e.*, slurry displacing water or chemical displacing slurry) in the pad-wafer region is critical for ensuring desired polish outcomes and enhancing module productivity. For instance, in applications where slurry is intended to replace water in the pad-wafer region, significant process transients may exist, which can manifest themselves in gradual increases in slurry concentration. Given that removal rate is strongly dependent on abrasive concentration, slow fluid replacement can significantly impact material removal. This problem is further exacerbated as the industry migrates to 300 mm wafers, because the transients associated with larger wafers will be more pronounced. This anticipated trend prompted the definition of a new dimensionless group referred to as TR, which was defined as the ratio of slurry MRT to polish time. This study analyzed the impact of TR on Preston's constant which dictated ILD removal rates. The polish time was held constant at 30 s, which is shorter than current typical polish times but is in line with polish times predicted to become mainstream within three to five years. Results showed that TR affected ILD removal. A comparison of concentration plots and numerical summations of removal rates throughout the polish was performed for three conditions with differing MRTs. These plots indicated that the discrepancy in oxide removal was due to the varying lengths of the abrasive concentration transient period. Experimental results indicated that at conditions corresponding to shorter MRTs, more oxide was removed than at conditions corresponding to longer MRTs. This trend was also predicted by the numerical summation of the concentration curve. Finally, is was shown that Preston's constant underwent a general decrease with TR. This was significant because at higher MRTs, the deviation between actual oxide removed and that predicted by Preston's equation was large. Because MRT affected oxide removal, and slurry flow rate, wafer pressure, and relative pad-wafer velocity in turn affected MRT, the choice of these parameters was shown to be critical in maintaining a successful and predictable CMP process. In addition, technology dictates operating under the conditions that give the highest Preston's constant and the highest level of control over removal rate.

Acknowledgments

The authors express their gratitude to DuPont Air Products Nanomaterials for slurry donation, and to Rodel-Nitta Company for donation of pads. This work was financially supported by the NSF/SRC Engineering Research Center for Environmentally Benign Semiconductor Manufacturing.

The University of Arizona assisted in meeting the publication costs of this article.

References

1. D. Stein, D. Hetherington, M. Dugger, and T. Stout, *J. Electron. Mater.*, **25**, 1623 (1996).
2. K. Kim, S. Moon, and H. Jeong, in *Chemical Mechanical Polishing in IC Device Manufacturing III*, R. L. Opila, C. Reidsema Simpson, K. B. Sundaram, I. Ali, Y. A. Arimoto, and Y. Homma, Editors, PV 99-37, pp. 402-407, The Electrochemical Society Proceedings Series, Pennington, NJ (1999).
3. A. Sikder, F. Giglio, J. Wood, A. Kumar, and M. Anthony, *J. Electron. Mater.*, **30**, 1520 (2001).
4. *The International Technology Roadmap for Semiconductors 2002*, Semiconductor Industry Association, San Jose, CA (2002).
5. A. Philipossian and E. Mitchell, *Micro*, **20**, 85 (2002).
6. F. Preston, *J. Soc. Glass Technol.*, **11**, 214 (1927).
7. O. Levenspiel, *Chemical Reaction Engineering*, John Wiley & Sons, Inc., New York (1972).
8. G. Froment and K. Bischoff, *Chemical Reactor Analysis and Design*, John Wiley & Sons, Inc., New York (1979).
9. E. Mitchell, M.S. Thesis, University of Arizona, Tucson, AZ (2002).